Ph.D. Dissertation

Polar Codes for Non-identically Distributed Channels and their Applications to Index Codes

비동일 분포 병렬 채널을 위한 폴라 부호 기법과 인덱스 코드를 위한 연계 폴라 부호 설계 기법

By

Jangseob Kim

August 2015

Department of Electrical Engineering and Computer Science
College of Engineering
Seoul National University

Abstract

In part I, we introduce new Polar coding schemes for independent non-identically distributed parallel binary discrete memoryless channels. The first scheme is developed for the case where underlying channels are time invariant (the case of deterministic channel parameter), and the other schemes deal with a scenario where underlying channels change based on a distribution (the case of random channel parameter). In the latter case, we model the channel behavior of binary erasure channels by exploiting a random variable. The proposed Polar coding schemes is shown to achieve the symmetric capacity.

We proved that for deterministic CPs in non-identical channel models, polar codes can achieve the sample mean of bit channel capacities. The key contribution is a new system model where the transmitter and the receiver knows only the channel parameter distribution instead of channel parameter itself. Though the existence can be proved by the mean value theorem on symmetric capacity I, it is not discussed how to find them. One may use an inverse function I^{-1} (or an approximate version) or pre-calculated table look-up. However, if the underlying channel type is BEC, the coding scheme can become simpler. Note that for a BEC with erasure probability ϵ, its symmetric capacity I is the affine function of ϵ. Then, we have the relation $E[I(\varepsilon)] = I(\overline{\epsilon})$ where $\overline{\epsilon}$ is the expectation of the random variable $\varepsilon \sim f_{\varepsilon}(\epsilon)$.

By applying multiple streams of polar codewords, we prove that the average capacity

of any B-DMCs under our scenarios is achievable. However, this is obtained by sacrificing the latency and complexity, since they stack multiple blocks during encoding and decoding process. Hence, these schemes might not be suitable in the systems where low latency or low complexity is required. Rather, it is more practical in storage systems such as flash memory devices where throughput is much important than latency. Especially, for flash memories, statistical responses such as a voltage threshold would change with time and with the number of accesses to a cell block. Hence, as the storage capacity increases, it is inefficient for a storage controller, to figure out exact states of every blocks or cells. If statistics on their changes are given instead, we can manage cells more efficiently using the proposed polar coding scheme. In addition, in the case of parallel channels where there exist statistically different random disturbances across channels, it would be difficult to track all the channel parameters. However, if their statistics are known to the transmitter and the receiver, we can deliver data up to the average capacity through polar codes by sacrificing latency. In such cases, polar codes are a promising option which maximizes the throughput.

Under the non-independent channel scenario, we assume that N transmit channels are grouped into channels with size r which is a power of two, so that we can deal with the scenario as a non-binary system. If N is not divisible by r ($N \mod r \neq 0$), puncturing may be used to fit the system into a q-ary system. The proposed polar codes appear to be promising for applications where only the knowledge of channel parameter distribution is available, and can be practical for storage applications such as flash memory devices.

In part II, we develop a joint coding scheme of Index codes and nested Polar codes (ICPC) under non-identical B-DMCs. In Chapter 5, we developed the joint coding scheme of nested Polar codes and the index codes which is denoted as ICPC schemes and proved that via ICPC the system can achieve the rate of $\frac{1}{r}\sum_{j=1}^{L} I(W_j)$.

In full SI case, I proved that there exist ICPC schemes w.p.1, and for arbitrary SI patterns we showed in Theorem 5 that ICPC schemes that achieve such rate would exist only when there are at least one feasible IC solution for each SI pattern and M. In addition, we also consider probabilistic SI where those information might be feedbacked from receivers or S estimates existence of messages in each receiver. We model the SI graph as a random digraph $\bar{G}(L, p)$ where edge connection probabilities are all identical to p, and suggest the upperbound of the expected rate would be $o(\sqrt{L})$ when there exist at least one feasible IC solution for ICPC shcemes, exploiting the minimum rank of a random graph result in the previous literature and Theorem 5.

Keywords: (Nested) Polar codes, non-identical channels, Index codes, ICPC.

Student number: 2010-30978

Contents

List of Figures

List of Tables

Part I

Polar codes for Non i.i.d. Parallel channels

Chapter 1

Introduction

1.1 Backgrounds

Polar codes which was introduced by Arikan [1] is a coding scheme that achieves the symmetric capacity of any discrete memoryless channel (DMC) by exploiting channel polarization phenomena. The block error rate converges to zero as the code length N goes to infinity.

One disadvantage of Polar codes is its dependency on the underlying channel. If the true channel is unknown to an encoder, the code rate can get larger or smaller than the capacity. In the former case, communication becomes unreliable and in the latter, the spectrum is not fully utilized. When the underlying channel is unknown to a decoder then it would produce erroneous estimations compared to the original, since the code design is based on a different channel. This performance gap was analyzed in [6], and proposed a robust approximation

technique for this mismatch. In [7], it was proved that polarization is possible even under a non-stationary case.

In this paper, we prove the achievability of Polar codes in the parallel channels model using the martingale processes and extend the model to where the channel parameters are no more fixed but exhibit some random behaviors. In addition, we briefly discuss the importance of the usage of the proper interleaver Q in the later section.

The communication scenario that the transmitter and the receiver do not know the channel parameter, but knows its domain (set) to which it belongs is known as the compound channel scenario [9]. The authors defined the compound capacity as the rate which one can reliably transmit data without knowing the underlying channel parameter. In [10] and [11], Polar codes that achieve the compound capacity are proposed. In the former, the unknown channel is deterministic during a codeword transmission and in the latter, the authors dealt with a deterministic compound parallel channels model.

Parallel channels are used to model statistical behaviors of communications in conveying multiple data simultaneously between nodes, connected with multiple links. The internal structure of storage systems or singular value decomposed MIMO channels are one of the examples of parallel channel models. In these channels, each link needs not be the same with each other, and their statistical behaviors may be different to each other. Then natural question arises whether the parallel channel capacity is achievable via Polar codes. Hof et. al. in [4, 5] proved that when all channel parameters (CP) are given, it is achievable by using

Polar codes under a non-identically distributed channel model by dividing the problem into degraded and non-degraded channel cases.

1.2 Scope and Organization

In this thesis, we show that polarization also occurs in the parallel channels, and provide a proof of achievability of Polar codes for general binary discrete memoryless channels (B-DMC) under a non-identically distributed assumption where deterministic CPs are given. Second, in contrast to previous literature, we also consider that CPs which describe underlying BECs are realizations of some random variables, and only their probability distributions are known to the transmitter and the receiver, instead of exact CPs.

It is similar to the compound channel scenario in that the transmitter does not know CPs, but different in that the CPs which are realization of random variables may not be the same with each other. The system model is depicted in Fig. 4.8. The role of the interleaver in this figure is discussed in Section 2.2.

Developing Polar codes for the randomly varying channels scenario is one of the noticeable issues. N binary erasure channels those which are fully described by the set of corresponding erasure probabilities $\epsilon_i, \forall i \in [1, N]$. For the i^{th} transmit channel at time instant η, consider the case where the parameter $\epsilon_i(\eta)$ becomes a realization of a random variable ε which follows a distribution f_ε. In this case, the encoder and the decoder do not know the exact $\epsilon_i(\eta)$ of the underlying channel W, thus existing Polar coding schemes would

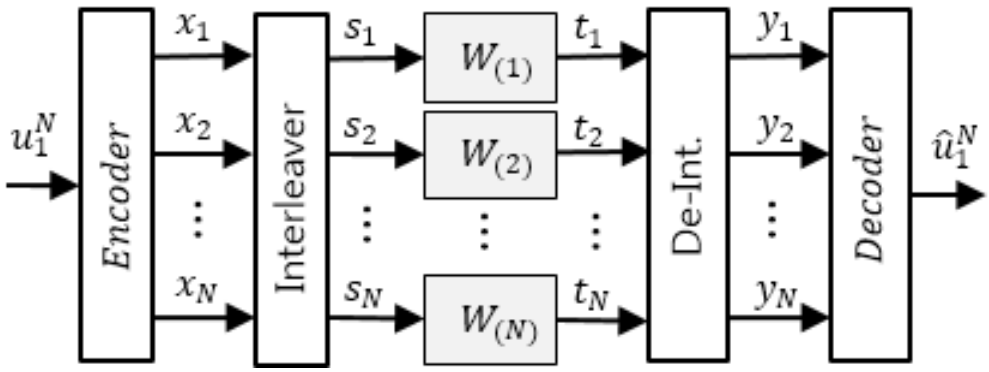

Figure 1.1 System model: Non-identically distributed parallel channels. Interleaver Q is inserted to make a set of virtually ordered transmit channels from $\{W_{(j)}\}$ to $\{W'_{(j)}\}, j \in [1, N]$. By applying the interleaver, $\{W_{(j)}\}$ is sorted in a way that maximize the reliability and the achievable rate.

not work.

For instance, in flash memories, the statistical responses such as a voltage threshold would change asymmetrically with time and with the number of accesses to cells. As the storage capacity increases, it is inefficient for a storage controller to probe exact states of every blocks or cells.

The rest of this paper is organized as follows. In Chapter 2, we analyze the polarization process on a non-identical parallel channel which is time invariant. This corresponds to a parallel channel with deterministic CPs which could be different with each other.

We also consider scenarios for a non-identically distributed channels with random CPs. We assume that each channel is independent with each other. In these sections, we assume that only distributions of CPs of BECs $W_{(i)}$ for $\forall i \in [1, N]$ are known to the transmitter and the receiver. In addition we consider the importance of the channel interleaver to enhance

the reliability of the SC-decoder.

In Chapter 3, CPs may change block by block manner, however, and they are constant within a block for all channels. Also, contrast to the previous section, we will let CPs may change even within a block.

Lastly, we shortly discuss a special case of partially dependent channels, and show that Polar codes based on q-ary input Polar codes would achieve the capacity, and concluding remarks are provided.

Chapter 2

Polar codes with deterministic non-identically distributed channels

2.1 Non-identical channels with deterministic CP

Let $W : \mathcal{X} \to \mathcal{Y}$ to denote a general symmetric binary input memoryless channels (B-DMC) and $W_N : \mathcal{X}^N \to \mathcal{Y}^N$ to denote a vector channel. If channels are independent but not identical, then $W_N(y_1^N|x_1^N) = \prod_{i=1}^N W_{(i)}(y_i|x_i)$ where $W_{(i)} : \mathcal{X}_i \to \mathcal{Y}_i$, such that their transition probabilities $p_{(i)}(y|x)$ and $p_{(j)}(y|x)$ may be different if $i \neq j$. In parallel channels, it could be a transmission through links with different qualities. In terms of channel model, this can interpreted as a fast fading channel in time or frequency domain. This corresponds to the case when the time duration or the frequency gap between adjacent channels is larger than coherence time or coherence frequency, respectively.

As in [1], given any B-DMC $W_{(i)}$, the same definitions of the symmetric capacity and

the Bhattacharyya parameter are adopted as performance measures:

- Symmetric capacity

$$I(W_{(i)}) \triangleq \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} \frac{1}{2} W_{(i)}(y|x) log \frac{W_{(i)}(y|x)}{\frac{1}{2}W_{(i)}(y|0) + \frac{1}{2}W_{(i)}(y|1)} \tag{2.1}$$

- Bhattacharyya parameter

$$Z(W_{(i)}) \triangleq \sum_{y_i\in\mathcal{Y}} \sqrt{W_{(i)}(y_i|0)W_{(i)}(y_i|1)}. \tag{2.2}$$

According to Proposition 1 in [1], the two parameters satisfy

$$log\frac{2}{1+Z(W_{(i)})} \leq I(W_{(i)}) \leq \sqrt{1-Z(W_{(i)})^2}. \tag{2.3}$$

For later use, we denote the bit channel $W_N^{(i)}$ by

$$W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) \triangleq \sum_{u_{i+1}^N\in\mathcal{X}^{N-i}} \frac{1}{2^{N-1}} W_N(y_1^N|u_1^N)$$

where $N = 2^n$ is the code length.

Now let us begin our discussion. In this section, channels are assumed to be independent and not identically distributed, and CPs are known to the transmitter and the receiver. For binary erasure channels (BEC) with ϵ_1^N, these erasure probabilities may be different with each other, and are known in advance to the encoder and the decoder.

Let us define the sum symmetric capacity by $I_s = \sum_1^N I_i$ where $I_i = I(X_i; Y_i)$, and sample mean by $E[I_i] = \frac{I_s}{N}$. Then, the following theorem holds.

Theorem 1. *For any set of B-DMCs* $\{W_{(j)}\}, j \in [1, N]$*, for arbitrary small* $\delta \leq 0$*, there exist polar codes that achieve sum capacity* I_s*, in the sense that as* N *which is power of two goes to infinity, the fraction of indices* $i \in [1, N]$ *satisfies:*

$$\frac{|\{i|I(W_N^{(i)}) \in (1-\delta, 1]\}|}{N} \to \frac{I_s}{N}$$
$$\frac{|\{i|I(W_N^{(i)}) \in [0, \delta)\}|}{N} \to 1 - \frac{I_s}{N}$$

For simple notations, denote $Z_N^{(i)} \triangleq Z(W_N^{(i)})$ and $I_N^{(i)} \triangleq I(W_N^{(i)})$. To prove the above theorem, first we have to clarify recursive structures of $W_N^{(i)}$, $Z_N^{(i)}$, and $I_N^{(i)}$. Second, we also need to prove that values of $I_N^{(i)}$ and $Z_N^{(i)}$ would converge to $\{0, 1\}$ as the code length N increases. For the second proof, the martingale convergence theorem and (2.3) will be used.

For a kernel $F = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$ of the generator matrix $G_N = B_N F^{\otimes n}$, the recursively evolving structure of $(W_N^{(i)}, W_N^{(i)}) \mapsto (W_{2N}^{(2i-1)}, W_{2N}^{(2i)})$ is the similar to that of recursive equation in [1] except for last recursions for the length 2.

Lemma 1. *Suppose* $(W_N^{(i)}, W_N^{(i)}) \mapsto (W_{2N}^{(2i-1)}, W_{2N}^{(2i)})$ *for some set of binary input channels. Then*

$$W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-1}|u_{2i}) = \sum_{u_{2i}} \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2}|u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2}|u_{2i})$$

$$W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1}|u_{2i}) = \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2}|u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2}|u_{2i})$$

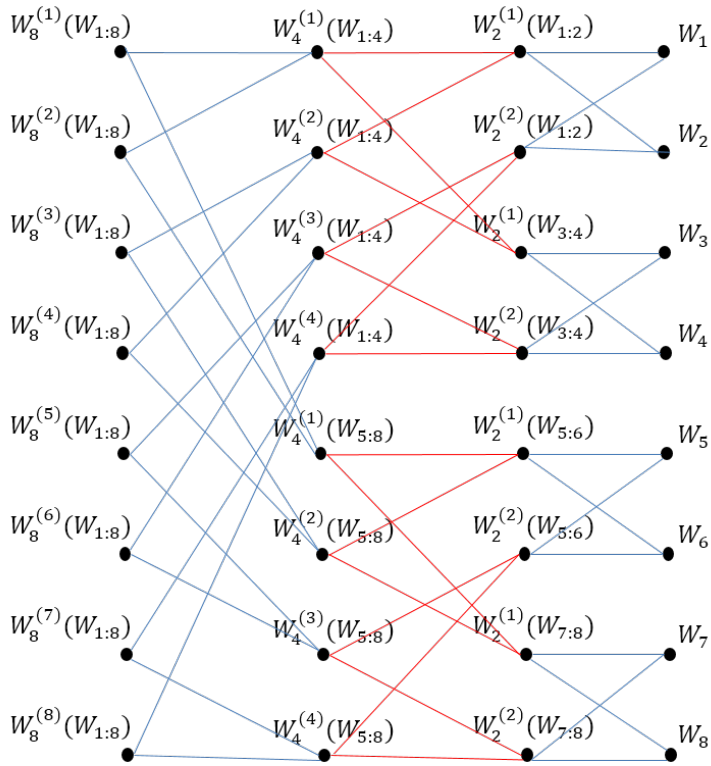

Figure 2.1 Recursive Parallel channel transitions $N = 8$. Note that the number of inputs and the outputs of each evolution is identical.

where the last recursive channel relations are mapping such that $(W_{(j)}, W_{(j+1)}) \mapsto (W_2^{(1)}, W_2^{(2)})$

$$W_2^{(1)}(y_j^{j+1}|u_j) = \sum_{u_{j+1}} \frac{1}{2} W_{(j)}(y_j|u_j \oplus u_{j+1}) \cdot W_{(j+1)}(y_{j+1}|u_{j+1}) \tag{2.4}$$

$$W_2^{(2)}(y_j^{j+1}, u_j|u_{j+1}) = \frac{1}{2} W_{(j)}(y_j|u_j \oplus u_{j+1}) \cdot W_{(j+1)}(y_{j+1}|u_{j+1}) \tag{2.5}$$

The evolution of these parallel channel transitions are depicted in Fig. 2.1. Observe that the structure is still recursive, but the number of inputs and the outputs of each evolution is identical.

The proof of the Lemma 1 is directly followed from [1]. The last recursion comes from the independently and non-identically distributed parallel channel environment.

Now we extend Proposition 4-7 in [1] which were proved under the i.i.d. condition, into the non-identically distributed DMC case. This extension has done in previous literatures, in the name of 'Parallel channels [5]' and 'Non-stationary channels [7]' with the measure of the symmetric capacity $I(\cdot)$.

2.1.1 The evolution of Symmetric Capacities

In this part, we consider the symmetric capacities of the recursively achieved bit-channels. By plugging (2.4) and (2.5) into the definition of the symmetric capacities (2.1), the following proposition holds.

Proposition 1. *Suppose* $(W_{(1)}, W_{(2)}) \mapsto (W_2^{(1)}, W_2^{(2)})$ *for any binary input discrete channels. Then*

$$I(W_2^{(1)}) + I(W_2^{(2)}) = I(W_{(1)}) + I(W_{(2)}) \tag{2.6}$$

$$I(W_2^{(1)}) \leq I(W_2^{(2)}) \tag{2.7}$$

$$I(W_2^{(2)}) \geq \max\{I(W_2^{(1)}), I(W_2^{(2)})\} \tag{2.8}$$

$$I(W_2^{(1)}) \leq \min\{I(W_2^{(1)}), I(W_2^{(2)})\} \tag{2.9}$$

Proof of (2.6). The equation (2.6) can be proved as follows:

$$\begin{aligned} I(W_2^{(1)}) + I(W_2^{(2)}) &= I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2, U_1) \\ &= I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2 | U_1) \\ &= I(U_1, U_2; Y_1, Y_2) \\ &= I(X_1, X_2; Y_1, Y_2) \\ &= I(X_1; Y_1) + I(X_2; Y_2) \\ &= I(W_{(1)}) + I(W_{(2)}) \end{aligned}$$

Since the mapping from a message vector U_1^2 to an encoded codeword is deterministic through a generator matrix G_2, there is no information loss between the third equation and the forth equation. And the fifth equation comes from the independence among transmit channels. □

The equation (2.7) obviously holds if (2.8) and (2.9) hold. To prove (2.7), focus on the second bit channel $W_2^{(2)}$

$$\begin{aligned} I(W_2^{(2)}) &= I(U_2; Y_1, Y_2, U_1) \\ &= I(U_2; Y_2) + I(U_2; Y_1, U_1 | Y_2) \\ &= I(X_2; Y_2) + I(U_2; Y_1, U_1 | Y_2) \\ &= I(W_{(2)}) + I(U_2; Y_1, U_1 | Y_2) \end{aligned}$$

$$\geq I(W_{(2)}) \tag{2.10}$$

since the value of mutual information is always non-negative, $I(\cdot) \geq 0$. By plugging (2.10) into (2.6) we get

$$I(W_2^{(1)}) \leq I(W_{(1)}) \tag{2.11}$$

However, it is yet ambiguous for (2.7) to be true based on (2.10) and (2.11), since there are no additional conditions on qualities of transmit channels $W_{(j)}, j \in [1, N]$ that would order them with measures of the symmetric capacity $I(\cdot)$ or the Bhattacharayya parameter $Z(\cdot)$.

Now we can consider three necessary conditions for (2.7) to be true:

1. $I(W_{(1)}) \leq I(W_2^{(2)})$
2. $I(W_{(2)}) \geq I(W_2^{(1)})$
3. $I(W_{(1)}) \leq I(W_{(2)})$

Conditions 1) and 2) are always true for for any BECs thus satisfy 2.7, however, for any B-DMC, it can not be assured whether it holds or not.

Then, If the third condition, $I(W_{(1)}) \leq I(W_{(2)})$, is satisfied in advance, (2.7) is satisfied without the need of previous two conditions and types of B-DMCs either. Note that transmit channels $\{W_{(i)} | i \in [1, N]\}$ are not aligned according to their qualities yet.

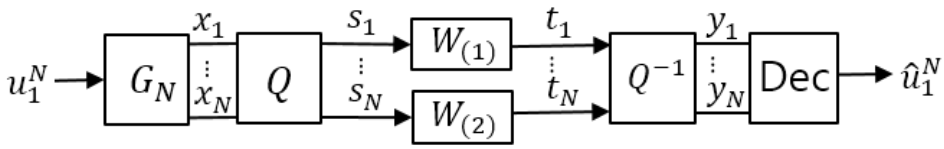

Figure 2.2 Interleaver Q is inserted to make a set of virtually ordered transmit channels from $\{W_{(j)}\}$ to $\{W'_{(j)}\}, j \in [1, N]$. By applying the interleaver, $\{W_{(j)}\}$ is sorted in a ascending order of the symmetric capacity and denoted as $\{W'_{(j)}\}$ such that $I(W'_{(l)}) \leq I(W'_{(m)})$ if $l < m$

Consider the fact that in this section, we have assumed the channel parameters are deterministic and already known to the transceiver (or the transmitter and the receiver can track qualities of all bit channels synchronously), hence we can virtually order the transmit channels based on their measured values by exploiting an interleaver Q at the transmitter and a de-interleaver Q' at the receiver. The transceiver structure including this interleaver and the de-interleaver is depicted in Fig. 2.10.

This interleaver would connect encoded bits to those properly ordered set of parallel channels that maximize the achievable rate. Therefore to find the appropriate Q can be views as an optimization problem.

In general, Q helps the polarization procedures by ordering transmit channels first.

For example, for $N = 4$, let the set of erasure probabilities of parallel BECs as $\{\epsilon_1^N\} = \{0.1, 0.4, 0.6, 0.9\}$ whose average is $\epsilon_m = 0.5$. The evolved result for this case is $\{I(W_N^{(i)})\} = \{0.02, 0.56, 0.44, 0.98\}$ and we choose the same indices as the information set: $A = \{2, 4\}$. The achievable amount of data of this interleaved one is 1.54 [bits] for two channel uses.

Now we apply the interleaver Q for the same set of $\{\epsilon_1^N\} = \{0.6, 0.4, 0.9, 0.1\}$. Then the resulting recursively evolved set of symmetric capacities is $\{I(W_N^{(i)})\} = \{0.02, 0.31, 0.69, 0.97\}$. The transmitter chooses two indices with highest $I(W_N^{(i)})$ values as the information set: $A = \{3, 4\}$, and the achievable sum is 1.66 [bits], which is 7.8% enhanced performance than before applying the interleaver.

Note that such an interleaver Q and its reverse Q^{-1} always exist for all code length N. Therefore, we can conclude that the relation (2.7) is true for all binary input DMCs, with the usage of appropriate interleavers of Q that would map the set of underlying transmit channels into an ordered set of virtual channels and could enhance the achievable rate. This complete the proof of (2.7) and the Proposition 1.

In general, we can extend these relations to the case of a code length $N = 2^n$. Then following relations are true for all given channel parameters.

$$I_{2N}^{(2i)} \leq min\{I_N^{(i)}(W_{[1:N]}), I_N^{(i)}(W_{[N+1:2N]})\} \tag{2.12}$$

$$I_{2N}^{(2i-1)} \geq max\{I_N^{(i)}(W_{[1:N]}), I_N^{(i)}(W_{[N+1:2N]})\} \tag{2.13}$$

$$I_{2N}^{(2i-1)} + I_{2N}^{(2i)} = I_N^{(i)}(W_{[1:N]}) + I_N^{(i)}(W_{[N+1:2N]}) \tag{2.14}$$

The equality holds between $I_{2N}^{(2i)}$ and $I_{2N}^{(2i-1)}$ if and only if underlying channels are either perfect or completely noisy. It is important to note that above inequalities hold.

According to equation (2.12)-(2.14), we can observe that the gap between those two evolved symmetric channel capacities $I_{2N}^{(2i-1)}$ and $I_{2N}^{(2i)}$ would increase as recursions are repeated (or as the code length N is doubled). In addition, their values are higher and lower that those of previous ones respectively.

Recalling that $I(\cdot) \in [0, 1]$, it might be estimated that those evolved $I_{2N}^{(2i-1)}$ and $I_{2N}^{(2i)}$ would converge to either 0 or 1, as the process goes on. This statement is proved via the Martingale convergence theorem next section.

2.1.1.1 Definition of the Martingale Process

For that, we introduce the definition of martingale process. Let (Ω, F, P) be a probability space and (T, τ) be a measurable space. And mostly consider $T = \mathbb{R}, \mathbb{R}^d, \mathbb{C}$. Unless otherwise indicated, it is to be understood from now on that $T = \mathbb{R}$.

Let $X = (X_n)_{n\geq 0}$ be a sequence of random variables taking values in T. We call X a stochastic process in T.

A filtration $(F_n)_n$ is an increasing family of sub algebras of F, i.e., $F_n \subset F_{n+1}$, for all n. We can think of F_n as the information available to us at time n. Every process has a natural filtration $(F_n^X)_n$, given by

$$F_n^X = \sigma(X_k, k \geq n)$$

The process X is called adapted to the filtration $(F_n)_n$, if X_n is $F_n - measurable$ for all n.

Of course, every process is adapted to its natural filtration. We say that X is integrable if X_n is integrable for all n.

Definition 1. *Let $(\Omega, F, (F_n)_{n\geq 0}, P)$ be a filtered probability space. Let $X = (X_n)_{n\geq 0}$ be an adapted integrable process taking values in $\mathbb{R}$.*

- *X is a martingale if $E[X_n|F_m] = X_m$ a.s., for all $n \geq m$*
- *X is a supermartingale if $E[X_n|F_m] \leq X_m$ a.s., for all $n \geq m$*
- *X is a submartingale if $E[X_n|F_m] \geq X_m$ a.s., for all $n \geq m$*

Note that every process which is a martingale (supermartingale, submartingale) with respect to the given filtration is also a martingale (supermartingale, submartingale) with respect to its natural filtration by the property of conditional expectation respectively.

2.1.1.2 Martingale process $\{I_n\}$

Following the notation of the random tree process as in [1], first we define the random sequence $\{I_n\}$ and $\{Z_n\}$ such that $I_N^{(i)} \mapsto I_n(b_1^n)$, where $I_n(b_1^n) \in \{I_n\}$ is represented as $I_{b_1^n}$, and $Z_N^{(i)} \mapsto Z_n(b_1^n)$, where $Z_n(b_1^n) \in \{Z_n\}$ is represented as $Z_{b_1^n}$ respectively.

Based on channel transitions in Fig. 2.1 and the definition of the average sequence, we can draw the alternative graphical transitions for each stage n as depicted in Fig. 2.3.

Then the following lemma shows that $\{I_n\}$ is a martingale process.

Lemma 2. *$\{I_n\}$ is a martingale under average sequence $\{E_n\}$ in the sense that $\mathbb{E}[I_{n+1}|E_n] = E_n$ where $I_{b_1^n} \in \{I_n\}$ and $E_n(b_1^n) = \mathbb{E}[I_{b_1^n}]$*

Proof. Lemma 2 can be proved from the chain rule of mutual information that preserves rate.

$$I_{2N}^{(2i-1)} + I_{2N}^{(2i-1)} = I_N^{(i)}(W_{[1:N]}) + I_N^{(i)}(W_{[N+1:2N]}) \tag{2.15}$$

By taking average on both sides, one can get

$$\mathbb{E}[I_n|b_1^n] = \frac{1}{2}((I_n(b_1^n, W_{[1:N]}) + I_n(b_1^n, W_{[N+1:2N]})) \tag{2.16}$$

which is equivalent to $E_n(b_1^n)$. Therefore, one can conclude that sequence $\{I_n\}$ is a martingale. □

Since $\{I_n\}$ is a bounded martingale, it converges to a random variable I_∞ with probability 1. Furthermore, from martingale property, $E[I_\infty] = E[I_0]$ which is equivalent to $\frac{I_s}{N}$. From (2.3), it follows that $I_\infty = 1 - Z_\infty$ for underlying BI-DMCs are BECs with non-identical erasure probabilities $\{\epsilon_1 \cdots \epsilon_N\}$. By combining this relation with the martingale property of the random sequence $\{I_n\}$, we have

$$P(I_\infty = 0) = 1 - \frac{I_s}{N} \tag{2.17}$$

$$P(I_\infty = 1) = \frac{I_s}{N}, \tag{2.18}$$

which concludes the proof of Lemma 3.

2.1.1.3 Example on BECs

Consider BECs with ϵ_1^N $(N = 2^n)$ where ϵ_i may be different with each other and $N = 2^{10}$. Channel parameters that describe those BECs are erasure probabilities of ϵ_1^N, and we assume that they are chosen in the range of $[0.4, 0.5]$. As is known, the BEC has a convenient property such that I_N can be described in a recursive form as follows:

$$\begin{aligned} I_{2N}^{(2i-1)} &= 1 - Z_{2N}^{2i-1} \\ I_{2N}^{(2i)} &= 1 - Z_{2N}^{2i} \end{aligned} \tag{2.19}$$

where

$$\begin{aligned} Z_{2N}^{(2i-1)} &= f(Z_N^{(i)}(W_{[1:N]}), Z_N^{(i)}(W_{[N+1:2N]})) \\ Z_{2N}^{(2i)} &= g(Z_N^{(i)}(W_{[1:N]}), Z_N^{(i)}(W_{[N+1:2N]})) \end{aligned} \tag{2.20}$$

Here, $f(\alpha, \beta) = \alpha + \beta - \alpha\beta$, $g(\alpha, \beta) = \alpha\beta$, and $W_{[1:N]} = \{W_{(j)}\}_{j=1}^N$. Note that equalities in (2.19) and (2.20) hold only for BEC. Normally, for arbitrary BI-DMCs, $I_{2N}^{(2i-1)} \geq 1 - Z_{2N}^{2i-1}$.

Fig. 2.4 depicts the polarization phenomena in $I_N^{(i)}$. Observe that even under the non-identically distributed channel condition, symmetric capacities would be polarized into

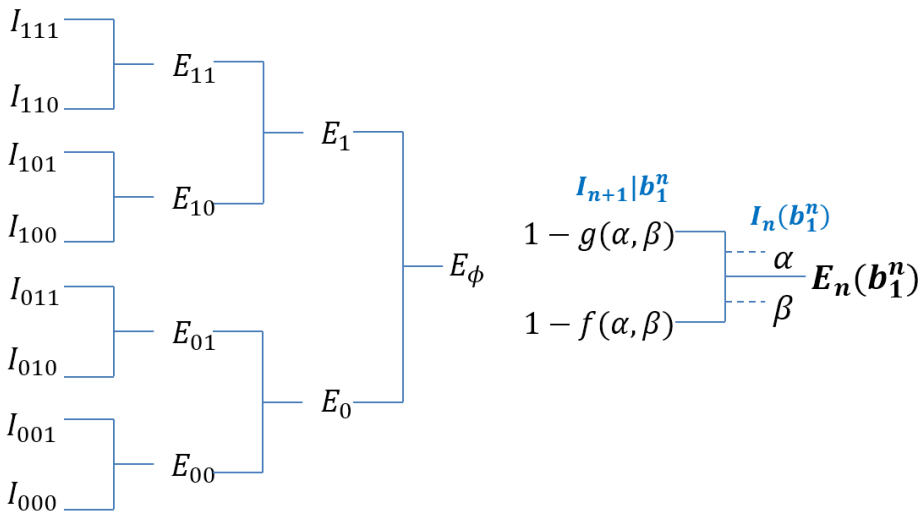

Figure 2.3 Graphical representation of the evolution of $I_N^{(i)}$ for $N = 8$ parallel BECs.

$\{0, 1\}$. It should be checked whether it approaches to the channel capacity as it is in i.i.d. In Table 2.1, one can check the difference between the symmetric capacity and the rate according to the exponent of n. Here, a difference is defined as $\frac{I_s - NR}{I_s}$. As we can see, the difference gets smaller as n increases, which means the rate R approaches to the average symmetric capacity $\frac{I_s}{N}$. Equivalently, it can be said that I_s is achievable in the sum rate sense.

Table 2.1 Difference between Capacities and Achievable Rates

n	8	10	12	14	16
Diff. [%]	0.6	0.044	0.038	0.003	0.002

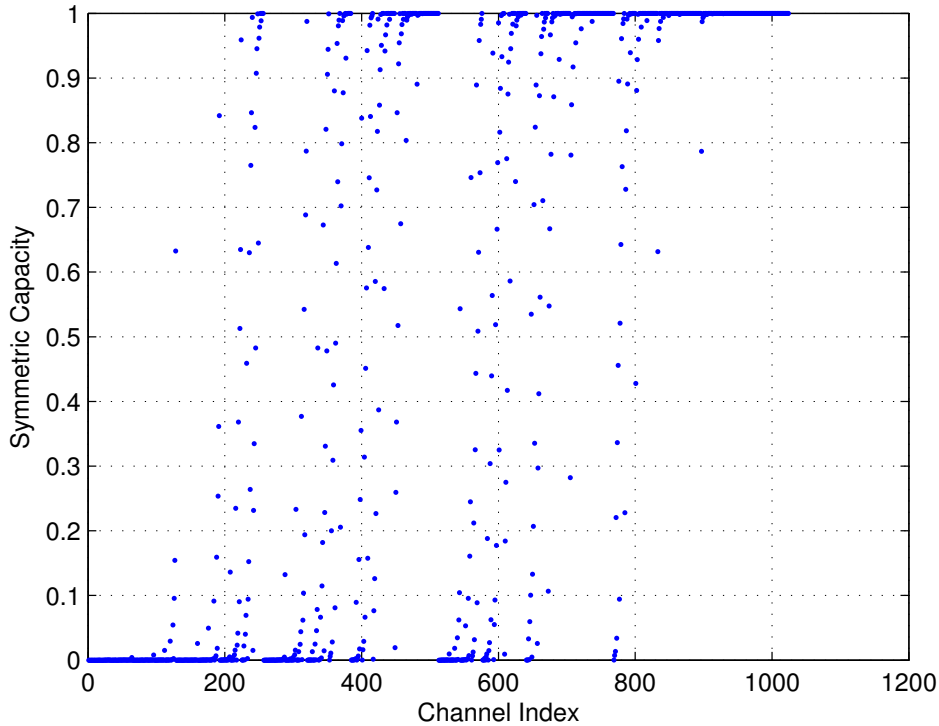

Figure 2.4 Plot of $I(W_N^{(i)})$ for a non-identical BEC(ϵ_1^N), $N = 2^{10}$.

2.1.2 Achievable Scheme based on the symmetric capacity

2.1.2.1 Encoder

Given CPs of $\{W_N^{(i)}\}$, the encoder calculates $\{I_N^{(i)}\}$ according to its definition. Only for BECs, $I_N^{(i)}$ calculation follows (2.19) with equality. In general BI-DMCs, one should track error performance of bit-channels with appropriate measures such as $Z_N^{(i)}$ by using the density evolution method [2]. Then, define an information index set $A = \{i | I_N^{(i)} \geq I_N^{(j)}, i \in A, j \notin A\}$ for all $i, j \in [1, N]$ such that $|A| = \lfloor I_s \rfloor$. Recall that we have to apply the interleaver Q

to map output codeword's indices to ordered transmit channels. G_N is an $N \times N$ generator matrix, and its kernel is $G_2 = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$. The encoder outputs a codeword $x_1^N = u_1^N \cdot G_N \cdot Q$.

2.1.2.2 Decoder

First the output sequence is provided to the successive cancellation (SC) decoder. Next, the receiver de-maps the output sequence according to the de-interleaver Q^{-1} and produces $\hat{u}_1^N$. One can easily check that it has the same recursive relations under the non-identical condition, by considering the evolution of bit channels.

$$L_N^{(i)}(y_1^N, u_1^{i-1}) = \frac{W_N^{(i)}(y_1^N, u_1^{i-1}|0)}{W_N^{(i)}(y_1^N, u_1^{i-1}|1)} \tag{2.21}$$

The only change is the last recursion of likelihood ratio (LR) equation: $L(y_i)$ changes to

$$L_i(y_i) = \frac{W_{(i)}(y_i|0)}{W_{(i)}(y_i|1)} \tag{2.22}$$

Thus, the decoding complexity is still maintained as $O(N \log N)$, and it has vanishing probability of error rate P_e as $N \to \infty$.

The encoding and decoding process is summarized in Algorithm 1.

2.1.3 The evolution of Bhattacharayya Parameters

Now we can show that the similar relationships as in the Proposition 1 on the Bhattacharayya functional $Z(\cdot)$.

Algorithm 1 Encoding and Decoding Process

Encoding Process

1: Given $\{W_{(i)}|1 \leq i \leq N\}$, calculate $\{I_N^{(i)}|1 \leq i \leq N\}$
2: Define A of size $k = I_s$ s.t. $A = \{i|I_N^{(i)} \geq I_N^{(j)}, \forall i \in A, \forall j \notin A\}$.
3: Fix u_F and insert data into $\{u_A|i \in A\}$.
4: Encode $x_1^N = u_1^N \cdot G_N \cdot Q$.

Decoding Process:

1: De-interleave: $y_1^N \cdot Q^{-1} \mapsto y_1^N$
2: Given (A, u_F), for $\forall i \in [1, N]$, calculate $L(y_1^N, u_1^{i-1})$ for u_i in the SC decoder based on modified LR in (2.21).

Proposition 2. *Suppose* $(W_{(1)}, W_{(2)}) \mapsto (W_2^{(1)}, W_2^{(2)})$ *for some binary input non-identically distributed discrete channels. Then following relations hold*

$$Z(W_2^{(1)}) \leq Z(W_{(1)}) + Z(W_{(2)}) - Z(W_{(1)})Z(W_{(2)}) \tag{2.23}$$

$$Z(W_2^{(2)}) = Z(W_{(1)})Z(W_{(2)}) \tag{2.24}$$

$$Z(W_2^{(2)}) \leq Z(W_2^{(1)}) \tag{2.25}$$

Proof of (2.24). The proof of this equation is straightforward.

$$\begin{aligned} Z(W_2^{(2)}) &= \sum_{y_1^2, u_1} \sqrt{W_2^{(2)}(y_1^2, u_1|u_2 = 0) W_2^{(2)}(y_1^2, u_1|u_2 = 1)} \\ &= \sum_{y_1^2, u_1} \frac{1}{2} \sqrt{W_{(1)}(y_1|u_1) W_{(2)}(y_2|0)} \cdot \sqrt{W_{(1)}(y_1|u_1 + 1) W_{(2)}(y_2|1)} \\ &= \sum_{y_2} \sqrt{W_{(2)}(y_2|0) W_{(2)}(y_2|1)} \sum_{y_1, u_1} \frac{1}{2} \sqrt{W_{(1)}(y_1|u_1) W_{(1)}(y_1|u_1 + 1)} \\ &= Z(W_{(2)}) Z(W_{(1)}) \end{aligned} \tag{2.26}$$

□

The fourth equation comes from that the summation over u_1 is a sum of two same terms.

In addition, the following relation hold with the aid of (2.24):

$$Z(W_2^{(2)}) \leq \min(Z(W_{(1)}), Z(W_{(2)})) \tag{2.27}$$

It can be verified simply by subtracting either of right terms from the left term.

Proof of (2.23). The proof of this equation is straightforward.

$$\begin{aligned}
Z(W_2^{(1)}) &= \sum_{y_1^2} \sqrt{W_2^{(1)}(y_1^2|u_1=0)W_2^{(1)}(y_1^2|u_1=1)} \\
&= \sum_{y_1^2} \sqrt{\sum_{u_2} \frac{1}{2} W_{(1)}(y_1|u_2)W_{(2)}(y_2|u_2) \sum_{u_2'} \frac{1}{2} W_{(1)}(y_1|1+u_2')W_{(2)}(y_2|u_2')} \\
&= \sum_{y_1^2} \frac{1}{2} \sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|0) + W_{(1)}(y_1|1)W_{(2)}(y_2|1)} \\
&\qquad \cdot \sqrt{W_{(1)}(y_1|1)W_{(2)}(y_2|0) + W_{(1)}(y_1|0)W_{(2)}(y_2|1)} \\
&\leq \sum_{y_1^2} \frac{1}{2} [\sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|0)} + \sqrt{W_{(1)}(y_1|1)W_{(2)}(y_2|1)}] \\
&\qquad \cdot [\sqrt{W_{(1)}(y_1|1)W_{(2)}(y_2|0)} + \sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|1)}] \\
&\qquad - \sum_{y_1^2} \sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|0)W_{(1)}(y_1|1)W_{(2)}(y_2|1)}
\end{aligned} \tag{2.28}$$

where the inequality follows from the identity

$$\begin{aligned}
[\sqrt{(\alpha\beta+\delta\gamma)(\alpha\gamma+\delta\beta)}]^2 + 2\sqrt{\alpha\beta\delta\gamma}(\sqrt{\alpha}-\sqrt{\delta})^2(\sqrt{\beta}-\sqrt{\gamma})^2 \\
= [(\sqrt{\alpha\beta}+\sqrt{\delta\gamma})(\sqrt{\alpha\gamma}+\sqrt{\delta\beta}) - 2\sqrt{\alpha\beta\delta\gamma}]^2
\end{aligned} \tag{2.29}$$

Then (2.28) becomes

$$\begin{aligned}
&\frac{1}{2}\sum_{y_1^2}\Bigg(W_{(1)}(y_1|0)\sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)} \\
&\qquad +W_{(1)}(y_1|0)\sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)} \\
&\qquad +W_{(1)}(y_1|0)\sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)} \\
&\qquad +W_{(1)}(y_1|0)\sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)}\Bigg) \\
&-\sum_{y_1^2}\sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|0)W_{(1)}(y_1|1)W_{(2)}(y_2|1)} \\
&\qquad = Z(W_{(1)}) + Z(W_{(2)}) - Z(W_{(1)})Z(W_{(2)})
\end{aligned} \tag{2.30}$$

Therefore, $Z(W_2^{(1)}) \leq Z(W_{(1)}) + Z(W_{(2)}) - Z(W_{(1)})Z(W_{(2)})$ is satisfied for any binary input channel parameters. □

Proof of (2.25). We can prove (2.25) simply by applying the arithmetic-geometric mean inequality on $Z(W_2^{(1)})$. Let us review the development process of (2.28):

$$\begin{aligned}
Z(W_2^{(1)}) &= \sum_{y_1^2}\sqrt{W_2^{(1)}(y_1^2|u_1=0)W_2^{(1)}(y_1^2|u_1=1)} \\
&= \sum_{y_1^2}\sqrt{\sum_{u_2}\frac{1}{2}W_{(1)}(y_1|u_2)W_{(2)}(y_2|u_2)\sum_{u_2'}\frac{1}{2}W_{(1)}(y_1|1+u_2')W_{(2)}(y_2|u_2')} \\
&= \sum_{y_1^2}\frac{1}{2}\sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|0)+W_{(1)}(y_1|1)W_{(2)}(y_2|1)} \\
&\qquad \cdot\sqrt{W_{(1)}(y_1|1)W_{(2)}(y_2|0)+W_{(1)}(y_1|0)W_{(2)}(y_2|1)}
\end{aligned} \tag{2.31}$$

Define shorthand notations of $A = W_{(1)}(y_1|0), B = W_{(1)}(y_1|1), C = W_{(2)}(y_2|0)$, and $D = W_{(2)}(y_2|1)$, we can rewrite above equation as follows

$$\begin{aligned} Z(W_2^{(1)}) &= \sum_{y_1^2} \frac{1}{2}\sqrt{ABC^2 + CDA^2 + CDB^2 + ABD^2} \\ &\geq \sum_{y_1^2} \frac{1}{2}\sqrt{4 \cdot ABCD} \\ &= \sum_{y_1} \sqrt{AB} \sum_{y_2} \sqrt{CD} \\ &= \sum_{y_1} \sqrt{W_{(1)}(y_1|0)W_{(1)}(y_1|1)} \sum_{y_2} \sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)} \\ &= Z(W_{(1)})Z(W_{(2)}) \\ &= Z(W_2^{(2)}) \end{aligned} \tag{2.32}$$

which the inequality is from the arithmetic and geometric mean relation of followings

$$ABC^2 + ABD^2 + CDA^2 + CDB^2 \geq 4 \cdot ABCD$$

□

Corollary 1. *Z_N can be described in a recursive form as follows:*

$$\begin{aligned} Z_{2N}^{(2i-1)} &\leq f(Z_N^{(i)}(W_{[1:N]}), Z_N^{(i)}(W_{[N+1:2N]})) \\ Z_{2N}^{(2i)} &= g(Z_N^{(i)}(W_{[1:N]}), Z_N^{(i)}(W_{[N+1:2N]})) \end{aligned} \tag{2.33}$$

where $f(\alpha,\beta)=\alpha+\beta-\alpha\beta, g(\alpha,\beta)=\alpha\beta$, *and* $W_{[1:N]}=\{W_{(j)}\}_{j=1}^{N}$. *The equality holds for BEC.*

By applying the recursive channel structure of Lemma 1 into the definition of the Bhattacharyaa parameter, the above corollary can be derived. that follows from [1].

With the aid of Lemma 1 and (2.23)-(2.25), we can derive Property 1, 2 and 3 as follows:

Property 1. $Z_{2N}^{(2i)} \leq min\{Z_N^{(i)}(W_{[1:N]}), Z_N^{(i)}(W_{[N+1:2N]})\}$

Property 2. $Z_{2N}^{(2i-1)} + Z_{2N}^{(2i)} \leq Z_N^{(i)}(W_{[1:N]}) + Z_N^{(i)}(W_{[N+1:2N]})$

Property 3. (for some BI-DMC W) $\quad Z_{2N}^{(2i-1)} \geq max\{Z_N^{(i)}(W_{[1:N]}), Z_N^{i}(W_{[N+1:2N]})\}$

Proof. Let us start from the first property. By using $f(\alpha,\beta)$ and $g(\alpha,\beta)$ notations as in the Corollay 1, the property 2) holds if and only if the joint event of $\{\alpha\beta \leq \alpha\} \cap \{\alpha\beta \leq \beta\}$ holds, which is true because of the non negativity of $\alpha, \beta \in [0,1]$ ($\alpha\beta - \alpha \leq 0, \alpha\beta - \beta \leq 0$).

The property 3) is able to be proved simply by adding two equation and inequality in (2.33) through the definition of $f(\alpha,\beta)$ and $g(\alpha,\beta)$.

The Property 1) can be derived from the definition of the parameter $Z(\cdot)$

$$Z(W_{(i)}) \triangleq \sum_{y_i \in \mathcal{Y}} \sqrt{W_{(i)}(y_i|0)W_{(i)}(y_i|1)}. \tag{2.34}$$

We use the same process in (2.32) that exploits the inequality relation between the arithmetic mean and the geometric mean, that will output $Z_{2N}^{2i-1} \geq Z_{2N}^{2i}$. □

We should note that, it is required that the Property 4) which is similar to (2.13), should be satisfied, to ensure the polarization phenomena of $Z(W_N^{(i)})$, with the bounded property $Z(W_N^{(i)}) \in [0, 1]$. However, it is generally not hold.

Property 3 holds in for some BI-DMCs such as Binary Erasure Channel, Binary Symmetric Channel and Binary input AWGN channel. To verify this, we look into the development of $Z_{2N}^{(2i-1)}$ for each channel type.

2.1.3.1 Additional Property for BEC

As mentioned, BEC has received attention for its special property that it maximizes the Z-parameter value of $f(\cdot)$ operation. For the BEC $W_{(i)}$ with erasure probability ϵ_i, $Z(W_{(i)}) = \epsilon_i$. For $N = 2$ with ϵ_1 and ϵ_2, $Z_2^{(1)} = \epsilon_1 + \epsilon_2 - \epsilon_1\epsilon_2$ and it is larger than ϵ_1 and ϵ_2. which means, by repeating recursions generally followings are satisfied

$$Z_{2N}^{(2i-1)} = Z_N^{(i)}(W_{[1:N]}) + Z_N^{(i)}(W_{[N+1:2N]})) - Z_N^{(i)}(W_{[1:N]}) \cdot Z_N^{(i)}(W_{[N+1:2N]}) \quad (2.35)$$

Hence, $\{Z_{2N}^{(2i-1)} \geq Z_N^{(i)}(W_{[1:N]})\}$ and $\{Z_{2N}^{(2i-1)} \geq Z_N^{i}(W_{[N+1:2N]})\}$, thus the Property 4) holds.

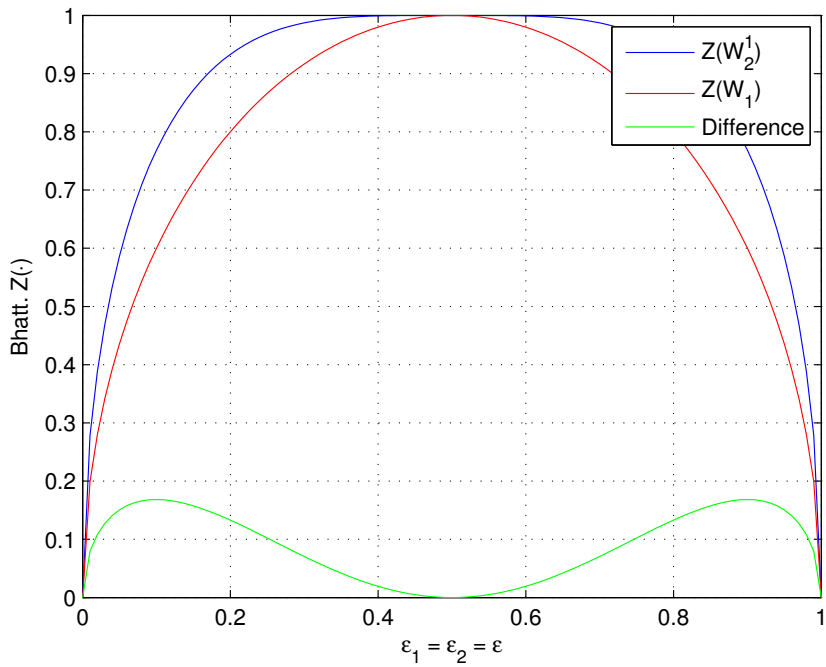

Figure 2.5 Difference in Z parameter for BSC $(N = 2)$ when $\epsilon_1 = \epsilon_2$

2.1.3.2 Additional Property for BSC

For $N = 2$ BSC with crossover probability ϵ_1 and ϵ_2, $Z(W_{(i)}) = 2\sqrt{\epsilon_i(1-\epsilon_i)}$. Then $Z_2^{(1)} = 2\sqrt{(1-\epsilon_s)\cdot\epsilon_s}$ where $\epsilon_s = \epsilon_1 + \epsilon_2 - 2\epsilon_1\epsilon_2$. Without loss of generality, assume that $\epsilon_1 \geq \epsilon_2$. If not, we swap these two channel using the interleaver Q. The numerical results in Fig. 2.5 and Fig. 2.6 depict that differences of $Z_2^{(1)} - Z(W_{(1)})$ for all CPs are always non-negative. Hence, we can conclude that the Property 3 holds for BSC case. The inequality (2.33) is also true for BSC, and it is shown in Fig. 2.7.

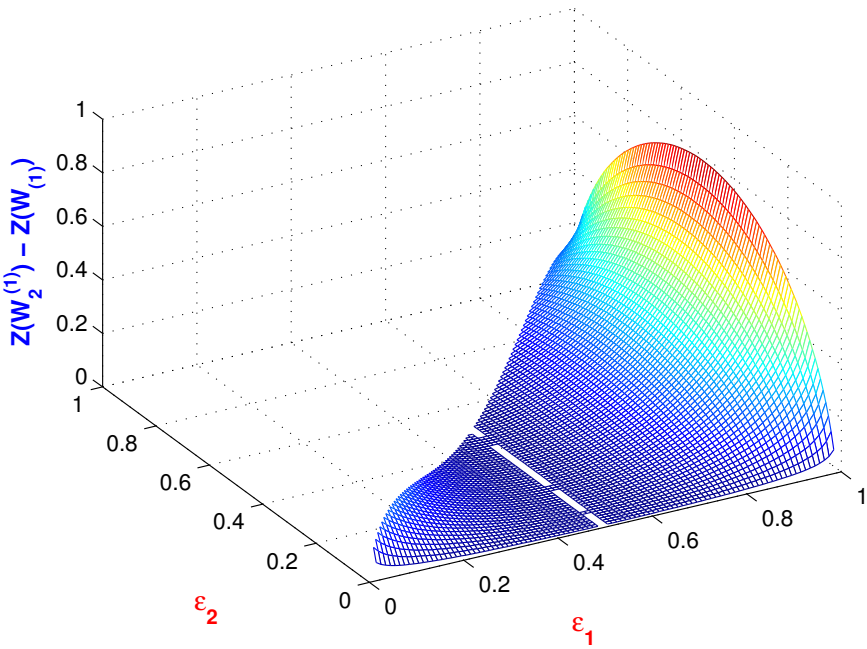

Figure 2.6 General Difference in Z parameter for BSC ($N = 2$) for all $\epsilon_1 \in [0, 1]$ and $\epsilon_2 \in [0, 1 - \epsilon_1]$. Domains come from the constraints $\epsilon_1 \geq \epsilon_2$.

2.1.3.3 Additional Property for BI-AWGN

Notice that binary input AWGN channels that follows the normal distribution $N(m, \sigma)$ can be transformed to equivalent BSC by applying a line coding that maps the set of binary alphabet $\{0, 1\} \mapsto \{-1, +1\}$. Therefore, since the Property 4) holds in BSC, it also holds in BI-AWGN channels.

Lemma 3. *For $i \in [1, N]$, $Z_N^{(i)}$ converges to {0,1} as $N \to \infty$, and the fraction of indices i converges to follows:*

$$\lim_{N \to \infty} \frac{|\{i | Z(W_N^{(i)}) \in (1 - \delta, 1]\}|}{N} = 1 - \frac{I_s}{N}$$

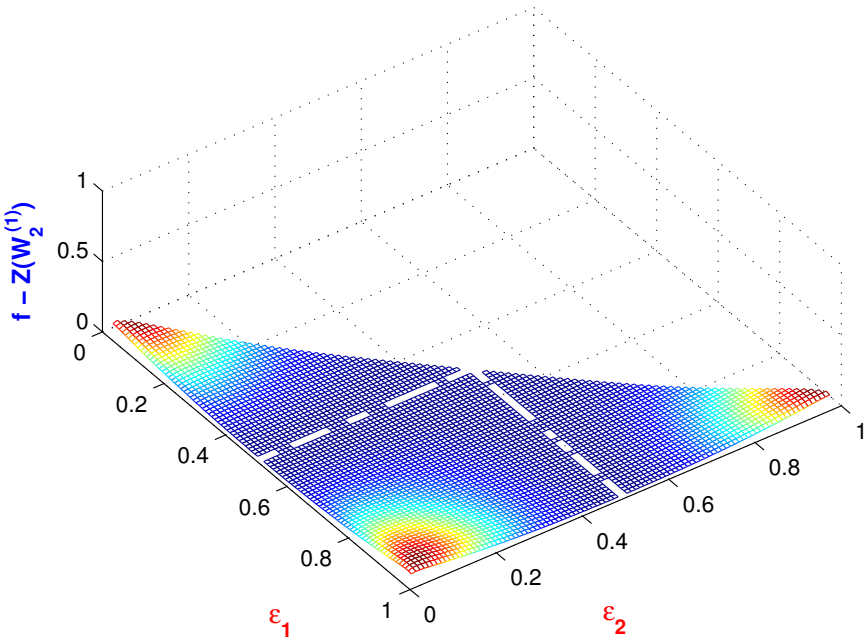

Figure 2.7 General Difference in Z parameter for BSC $(N = 2)$ for all $\epsilon_1 \in [0, 1]$ and $\epsilon_2 \in [0, 1 - \epsilon_1]$. Domains come from the constraints $\epsilon_1 \geq \epsilon_2$.

$$\lim_{N\to\infty} \frac{|\{i|Z(W_N^{(i)}) \in [0, \delta)\}|}{N} = \frac{I_s}{N}$$

Lemma 3 is an alternative version of the last condition in Theorem 1. As we know, the computational compexity in calculating Bhattcharyaa parameters are much less than that of symmetric capacities, therefore we focus on proving Lemma 3 to complete the proof of Theorem 1. For that, we show that the random sequence of Bhattacharyya parameters form a supermartingale, and its convergence property under the non-identical assumption.

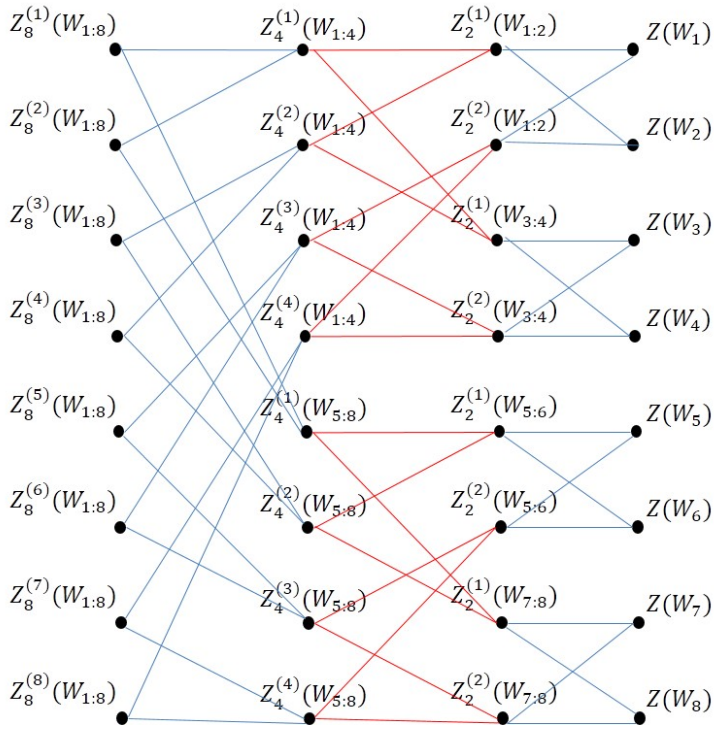

Figure 2.8 Graphical representation of the evolution of $Z_N^{(i)}$ for $N = 8$ parallel BECs.

2.1.4 Supermartingale Z_n

Following the notation of the random tree process as in [1], define the random sequence $\{I_n\}$ and $\{Z_n\}$ such that $I_N^{(i)} \mapsto I_n(b_1^n)$, where $I_n(b_1^n) \in \{I_n\}$ is represented as $I_{b_1^n}$, and $Z_N^{(i)} \mapsto Z_n(b_1^n)$, where $Z_n(b_1^n) \in \{Z_n\}$ is represented as $Z_{b_1^n}$ respectively.

Then for some BI-DMCs that satisfy four properties, we can draw a graphical evolving recursive structure of $Z_N^{(i)}$ (or Z_n). In Fig.2.8 we depict the graphical structure of Z_n for $N = 8$. This graph is similar to that of the identically distributed DMC case, however elements in the rightmost column of this figure are non-identically distributed, in a sense

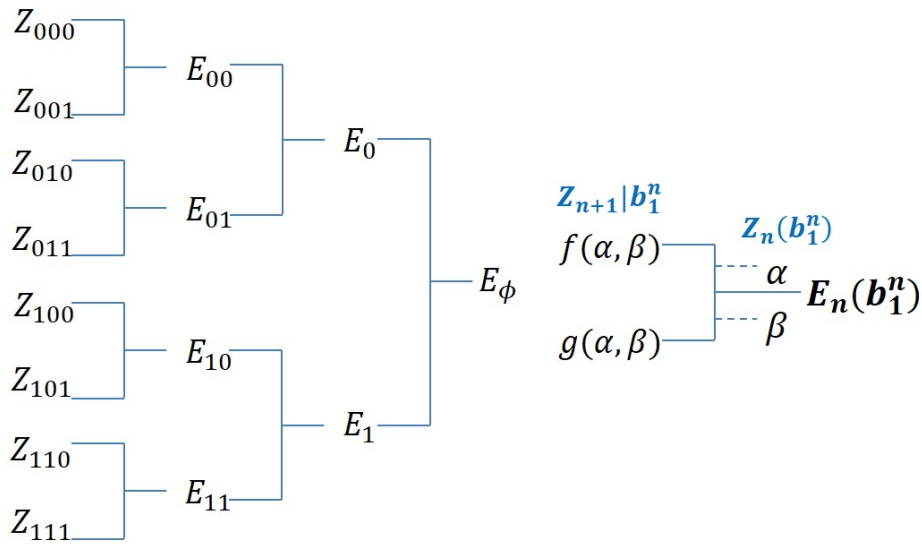

Figure 2.9 Alternative representation of $\{Z_n\}$ sequence for non-identically distributed parallel BECs.

that their channel parameters may not be the same.

$\{Z_n\}$ can be considered as a supermartingale if this random tree proess satisfies the relation $Z_n \geq \mathbb{E}[Z_{n+1}|b_1^n]$. Under the i.i.d. channel assumption where $W_{(i)} = W_{(j)} \forall i, j$, it is easily verified, since the one step transition from $Z_n(b_1^n)$ to $\{Z_{n+1}\}$ is a single variable to single variable mapping: $\mathcal{K} : \mathcal{Z} \rightarrow \mathcal{Z}$ such that $\mathcal{Z} = \{Z|Z \in \Re, Z \in [0,1]\}$. In contrast, with the non-identically distributed assumption where $W_{(i)}$ is not necessarily the same as $W_{(j)}$ when $i \neq j$, the transition becomes a two to one mapping: $\mathcal{K}' : \mathcal{Z}^2 \rightarrow \mathcal{Z}$. As shown in Fig.2.8, each $Z_{b_1^n}$ is not a scalar but a 2×1 vector. In this case, the format of the condition is not appropriate due to dimension mismatch. To resolve this mismatch, we introduce an

average sequence $\{E_n\}$ to replace $Z_{b_1^n}$ such that $E_n(b_1^n) = \mathbb{E}[Z_{b_1^n}]$. Then from (2.33), we can verify that $\mathbb{E}[Z_{n+1}|E_n(b_1^n)] \leq E_n$ which means Z_n is a supermartingale under E_n.

2.1.5 Convergence of $\{Z_n\}$

Since $\{Z_n\}$ is bounded within $[0, 1]$ and supermartingale, from the martingale convergence theorem, there exists a random variable Z_∞ that the sequence $\{Z_n\}$ converges with probability 1 as $n \to \infty$. It is equivalent to the statement that in L^1, $\lim_{n\to\infty} \mathbb{E}[|Z_{n+1} - E_n|] = 0$ which implies $E[|Z_\infty - E_\infty|] = 0$. Given that

$$E_n(b_1^n) = \frac{\alpha + \beta}{2}, Z_{n+1}|(b_1^n) = \begin{cases} f(\alpha, \beta) & w.p.\frac{1}{2} \\ g(\alpha, \beta) & w.p.\frac{1}{2} \end{cases}$$

where, $\alpha \triangleq Z_N^{(i)}(W_{1:N})$ and $\beta \triangleq Z_N^{(i)}(W_{N+1:2N})$, it becomes an indeterminate equation with the condition of $\alpha, \beta \in [0, 1]$. One can obtain solution pairs $(\alpha_\infty, \beta_\infty) = \{(0, 0), (1, 1)\}$. Therefore, since Z_∞, which corresponds to α_∞, would converge to either 0 or 1 almost everywhere, we can conclude that all $Z_N^{(i)}$ are polarized in either near perfect or totally random manner as $n \to \infty$.

2.2 Channel mapping via the Interleaver Q

According to the basic theory of Polar codes, channels are recursively evolved such that $(W_N^{(i)}, W_N^{(i)}) \mapsto (W_{2N}^{(2i-1)}, W_{2N}^{(2i)})$ for some set of binary input discrete memoryless channels.

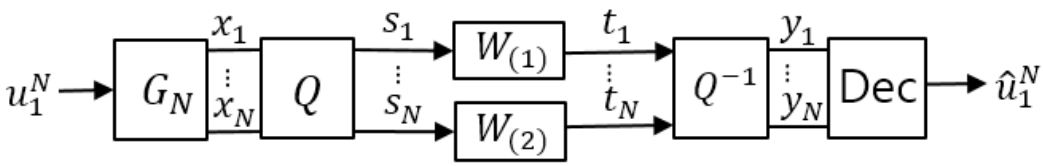

Figure 2.10 Interleaver Q is inserted to make a set of virtually ordered transmit channels from $\{W_{(j)}\}$ to $\{W'_{(j)}\}, j \in [1, N]$. By applying the interleaver, $\{W_{(j)}\}$ is sorted in a way that maximize the reliability and the achievable rate.

We denote each mapping as follows:

$$\mathcal{F}_{n+1} : W_N^{(i)}, W_N^{(i)} \mapsto W_{2N}^{(2i-1)} \tag{2.36}$$

$$\mathcal{G}_{n+1} : W_N^{(i)}, W_N^{(i)} \mapsto W_{2N}^{(2i)} \tag{2.37}$$

and it was proved that polarizations would occur in non-stationary channels.

In this section, we discuss the role of an interleaver Q in Polar coding systems under non-identically distributed B-DMCs, and propose an algorithm that explains how to construct such an operation.

The transceiver structure including an interleaver Q and a de-interleaver Q^{-1} is depicted in Fig. 2.2. To understand the importance of Q, let us consider the following example.

Example 1. *For $N = 4$, let the set of erasure probabilities of parallel BECs as $\{\epsilon_1^N\} = \{0.1, 0.4, 0.6, 0.9\}$ whose average is $\epsilon_m = 0.5$. Then the evolved bit channels capacities are $\{I(W_N^{(i)})\} = \{0.02, 0.56, 0.44, 0.98\}$ and the encoder chooses the information set $A = \{2, 4\}$. The resulting sum of chosen bit channels' capacities is 1.54 [bits] for four channel uses.*

Now we apply the interleaver Q over the same set which results in $\{\epsilon_1^N\} = \{0.6, 0.4, 0.9, 0.1\}$. *The evolved set of symmetric capacities is* $\{I(W_N^{(i)})\} = \{0.02, 0.31, 0.69, 0.97\}$.

The encoder chooses two indices with highest $I(W_N^{(i)})$ *values:* $A = \{3, 4\}$*, and the achievable sum is 1.66 [bits], which is* 7.8% *enhancement compared to the previous one.*

Note that the larger the sum, $\sum_{i \in A} I(W_N^{(i)})$, is, more reliable the system is. The result of this example indicates that for a given set of channels $\{W_{(i)} | i \in [1, N]\}$, there would be the optimal mapping $Q : x_1^N \mapsto s_1^N$ in a sense that maximizing the reliability (or the rate) by boosting polarizations among bit channels' symmetric capacities. Finding the appropriate Q can be views as an optimization problem:

$$\begin{aligned} Find \quad & Q : x_1^N \mapsto s_1^N \\ Maximize \quad & \sum_{i \in A} I(W_{(i)}) \end{aligned}$$

Equivalently, the mapping can be interpreted as the channel permutation such that $Q : \{W_{(i)}\} \mapsto \{W'_{(i)}\}$. We discuss two methods that search such a mapping Q.

2.2.1 Exhaustive Search Method with Grouping

The simplest yet naive approach is try every possible combinations over N channels and choose the best one.

Obviously, there are $N!$ number of cases to check. However, we can categorize every combinations into equivalent groups in a sense that in each group, all combinations output the same qualities of bit channels. The size of each group is 2^{N-1}. Hence, owing to the recursive channel evolving structure, the required number of tests is $\frac{N!}{2^{N-1}}$. The detailed proof is in the Appendix D.

This grouping technique considerably reduces the computational burdens: for $N = 8$, we need to test 315 representative combinations instead of $N! = 40320$. However, unfortunately the enhanced test set size would go beyond the computational capability for practical N lengths.

2.2.2 Heuristic method

The purpose of the channel combining and the splitting operation is to build virtual channels that are as close to the extremal channels as possible as recursions are repeated.

We already know that for any non-identical parallel B-DMCs, there exist Polar codes that achieve the symmetric capacity as N goes to the infinity. However, in a practical systems that exploit finite code lengths, we observed via the Example 1, that there would be differences in convergence speed for different wire-lings through the interleaver Q.

A well designed Q will polarize bit channel qualities fast. To that end, it should sort the given set of transmit channels in order to create as many enhanced and degraded those which are closer to the extremals simultaneously as possible. In this section, we propose an

algorithm of Q that achieves such an object.

1. Sort transmit channels $W_{(i)}, i \in [1, N]$ in an ascending order of the capacity $I(W_{(i)})$.

2. Make $\frac{N}{2}$ pairs: the i^{th} pair includes the i^{th} smallest one and the i^{th} largest one.

3. Using the indices of $\frac{N}{2}$ pairs, [1: $\frac{N}{2}$], repeat the second procedure until the size of index set becomes 4.

As can be seen, this algorithm has a recursive structure, and we can represent it in a matrix form. Let P_n represent the interleaver operation Q for length $N = 2^n$ and S_n indicate the 2nd operation of the above algorithm. Then

$$P_n = S_n \cdot P_{n-1}^{\otimes \mathcal{I}_2} \tag{2.38}$$

where $\mathcal{I}_2$ is the 2×2 identity matrix and $\otimes$ means the kronecker product.

The interleaving procedures are summarized in Algorithm 2.

Algorithm 2 Find Mapping Q

Initial $P_2 = S_2$
Sort transmit channels in the order of $I(W_{(i)})$
while $n \geq 2$ **do**
    Perform pairing S_n
    $P_n = S_n \cdot P_{n-1}^{\otimes \mathcal{I}_2}$
end while

In Fig. 2.11, we depict achievable rates $\frac{1}{N} \sum_{i \in A} I(W_N^{(i)})$ of Polar codes by exploiting different interleaving methods under parallel BECs whose erasure probabilities are uniformly chosen within the range of $(0, 1)$ (hence the average 0.5).

The black curve corresponds to the capacity, the red one is the result of the proposed algorithm and the blue one is the performance when using a random shuffling. And we can get the green curve when we apply the proposed algorithm only once without recursion: that is, $P_n = S_n$. Finally the scarlet curve is the performance when the transmit channels are sorted in capacities, ascending or descending order does not affect the result.

From the figure, the proposed interleaving algorithm shows better performance in the rate than the others. And it is observed that when $W_{(i)}$s forms ordered set, it would converge to the capacity slower than others. Hence if exposed channels' symmetric capacities are ordered in either ways, they should be rearranged through Q .

2.3 Link failures: Puncturing operation

Suppose that there are link failures in some transmit channels. This loss in channels should be reflected to the Polar coding structure such as the encoder and the decoder. Especially, those failures would affect the information set A since it is channel dependent,

The punctured position under the existence of non-identical parallel B-DMCs should also be cautiously dealt with. When under the i.i.d. channel case, to find puncturing positions among x_1^N, the authors in [16] suggested that let the punctured position's capacities be all zeros and find A.

However, in deterministic parallel non-identical channels case, it becomes simpler: to minimize the loss in achievable rate, we should puncture transmit channels with lowest

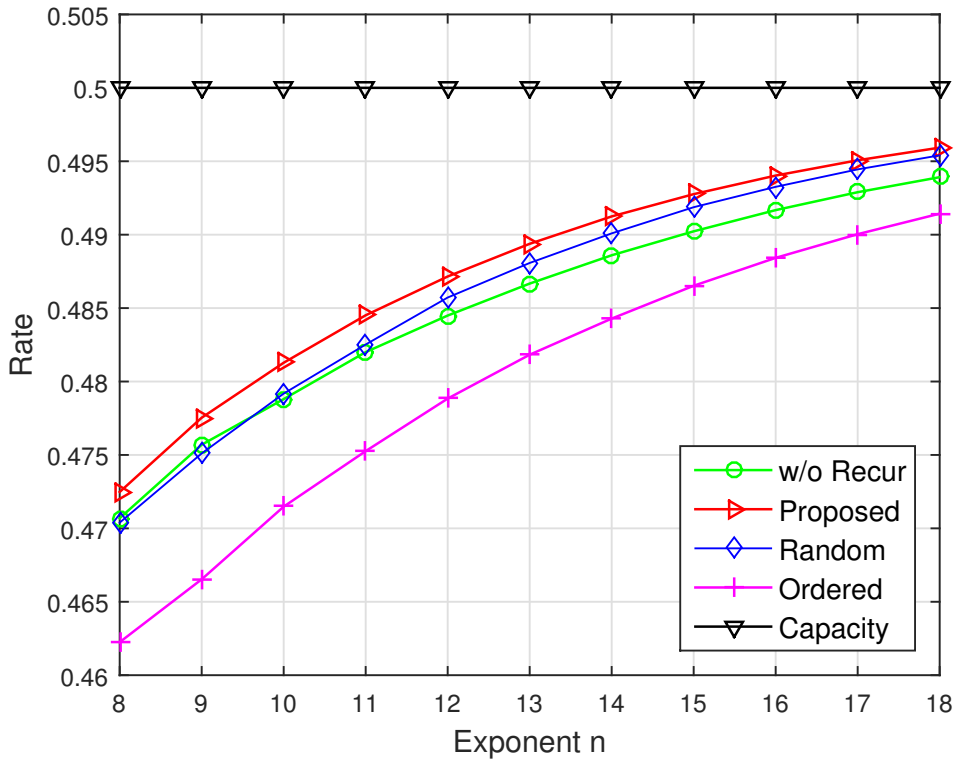

Figure 2.11 Observation on the different usage of the interleaver Q. Shuffling the coded alphabets x_1^N as randomly as possible exhibits higher achievable rates compared to the ordered interleaver.

symmetric capacities. Then we let their capacities are all zeros, and apply recursions to find A. Contrast to the previous puncturing systems, we consider the interleaver Q in defining A. Q will map channels with failures into the another combination to minimize the degraded convergence speed.

2.4 Polarizations on non-independent channels

In previous sections, it is assumed that all transmit channels are independent with each other in the sense that

$$W_N(y_1^N|x_1^N) = \prod_{i=1}^{N} W_{(i)}(y_i|x_i) \tag{2.39}$$

where $W_{(i)} : \mathcal{X}_i \to \mathcal{Y}_i$. Now consider the case where these channels have correlation among them, thus the equality in (2.39) does not hold. Then, the previous polar coding scheme for independent B-DMCs may not achieve the capacity.

As an example, check out the polarization phenomenon in a measure of the symmetric capacity for $N = 2$ where the transition probability of the channel $W_2 : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}_1 \times \mathcal{Y}_2$ where $\mathcal{X}_1$ and $\mathcal{X}_2$ are $GF(2)$, and $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1, e\}$ is defined as follows:

$$W_2(y_1^2|x_1^2) = \begin{cases} 1-\epsilon, & y_1^2 = x_1^2 \\ \\ \epsilon & y_1^2 = e \end{cases}$$

The element e denotes the erasure symbol. Without losing information, W_2 can be modelled as a single quaternary erasure channel (QEC) W' with CP ϵ whose capacity is $I(W') = 2(1-\epsilon)$.

Now apply the same generator matrix G_2 to the encoder, and calculate the symmetric capacities of split bit channels $I(W_2^{(1)})$ and $I(W_2^{(2)})$ via (2.1). Then we can get $I(W_2^{(1)}) = I(W_2^{(2)}) = 1 - \epsilon$ which means that no polarization occurs with the coding strategy for independent channels. By relabeling, the set of binary input vectors $X_1^2 = \{00, 01, 10, 11\}$ is

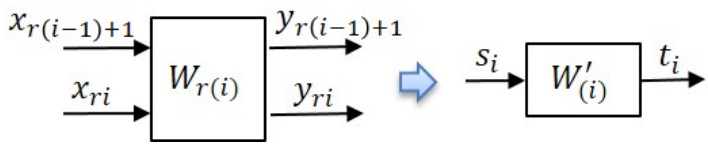

Figure 2.12 Relabeling of a length r=2 binary vector channel to a single quaternary channel.

mapped to a quaternary symbol set $S = \{0, 1, 2, 3\}$ as depicted in Fig. 2.12 where $r = 2$ with the channel notation change from the correlated vector channel W_2 to the single independent q-ary DMC W'. The problem of non-polarization is proved to occur in some q-ary DMCs when the cardinality of the set S is a composite number [12]. However, in [13] the authors proved the existence of polar codes for composite cardinality of $q = 2^r$.

Under the parallel channel model, when $N = 2^n$ transmit channels are pairwise correlated in the sense that

$$W_N(y_1^N|x_1^N) = \prod_{i=1}^{N/2} W_{2(i)}(y_{2i-1}^{2i}|x_{2i-1}^{2i}), \tag{2.40}$$

it can be represented as N' length of independent quaternary input DMCs:

$$W_{N'}(t_1^{N'}|s_1^{N'}) = \prod_{i=1}^{N'} W'_{(i)}(t_i|s_i) \tag{2.41}$$

where $N' = \frac{N}{2}$.

For generalization, suppose that r transmit channels are correlated in a similar sense of (2.40). If we assume $r = 2^\alpha$ $(1 \leq \alpha \ll n - 1)$ for simplicity, then the new code length becomes $N' = 2^{n-\alpha}$, and we have

$$W_N(y_1^N|x_1^N) = \prod_{i=1}^{N'} W'_{r(i)}(y_{r(i-1)+1}^{ri}|x_{r(i-1)+1}^{ri}). \tag{2.42}$$

In a q-ary representation, (2.41) still holds where the input alphabet cardinality is $q = 2^r$. If those relabeled independent q-ary DMCs are identically distributed, such that $W'_{(i)} = W'_{(j)}$ for $\forall i, j \in [1, N']$, it is proved in [13] that polar codes for q-ary input DMCs achieves the symmetric channel capacity when $q = 2^r$ by exploiting the same kernel as in [1].

Therefore, the following proposition holds for a general q-ary DMC W':

Proposition 3. *There exist polar codes for non-independent DMCs W' that achieve the symmetric $I(W')$, by relabeling r-bits binary sequence to a q-ary ($q = 2^r$) symbol as $N' \to \infty$ through power of two.*

Proof. The proof directly follows from [13], which proves the existence of polar codes which achieve the symmetric capacity for q-ary input DMCs when q is a power of two. Since the relabeling of input alphabet causes no information loss, the achievability still holds. □

2.5 Summary

In section II, we proved that for deterministic CPs in non-identical channel models, polar codes can achieve the sample mean of bit channel capacities. In Section III and IV, the key contribution is a new system model where the transmitter and the receiver knows only the channel parameter distribution instead of channel parameter itself. Though the existence can be proved by the mean value theorem on symmetric capacity I, it is not discussed how to find them. One may use an inverse function I^{-1} (or an approximate version) or pre-calculated table look-up. However, if the underlying channel type is BEC, the coding scheme can

become simpler. Note that for a BEC with erasure probability ϵ, its symmetric capacity I is the affine function of ϵ. Then, we have the relation $E[I(\varepsilon)] = I(\overline{\epsilon})$ where $\overline{\epsilon}$ is the expectation of the random variable $\varepsilon \sim f_{\varepsilon}(\epsilon)$.

By applying multiple streams of polar codewords, we prove that the average capacity of any B-DMCs under our scenarios is achievable. However, this is obtained by sacrificing the latency and complexity, since they stack multiple blocks during encoding and decoding process. Hence, these schemes might not be suitable in the systems where low latency or low complexity is required. Rather, it is more practical in storage systems such as flash memory devices where throughput is much important than latency. Especially, for flash memories, statistical responses such as a voltage threshold would change with time and with the number of accesses to a cell block. Hence, as the storage capacity increases, it is inefficient for a storage controller, to figure out exact states of every blocks or cells. If statistics on their changes are given instead, we can manage cells more efficiently using the proposed polar coding scheme. In addition, in the case of parallel channels where there exist statistically different random disturbances across channels, it would be difficult to track all the channel parameters. However, if their statistics are known to the transmitter and the receiver, we can deliver data up to the average capacity through polar codes by sacrificing latency. In such cases, polar codes are a promising option which maximizes the throughput.

Under the non-independent channel scenario, we assume that N transmit channels are grouped into channels with size r which is a power of two, so that we can deal with the

scenario as a non-binary system. If N is not divisible by r ($N \mod r \neq 0$), puncturing may be used to fit the system into a q-ary system. The proposed polar codes appear to be promising for applications where only the knowledge of channel parameter distribution is available, and can be practical for storage applications such as flash memory devices.

In addition, we also proved that for deterministic CPs in non-identical channel models, polar codes can achieve the sample mean of bit channel capacities. In this chapter, we show that polarization also occurs in the parallel channels, and provide a proof of achievability of polar codes that achieve the sample mean of bit channel capacities. for general binary discrete memoryless channels (B-DMCs) under a non-identically distributed assumption where deterministic CPs are given.

Chapter 3

Non-identical Binary Erasure Channels with random Erasure probabilities with Single distribution

3.1 Non-identical Binary Erasure Channels with random Erasure probabilities with Single distribution

In previous literatures [5] and [7], it is assumed that underlying discrete memoryless channels' characteristics are fully exposed to the transceiver thus encoder and the decoder exploit those information. Under this condition, it is proved that Polar codes can achieve the symmetric capacity.

In this section, we assume that channel parameters are not deterministic but realizations of a random variable. For BEC, the channel transition probability of a transmit channel $W_{(i)} : \mathcal{X}_i \mapsto \mathcal{Y}_i$ is fully described by the erasure probability ϵ_i. Hence channel features of the non-identically distributed parallel channels model would be perfectly represented via

the set of erasure probabilities $\{\epsilon_1^N\}$.

Random erasure probabilities means that each ϵ_i is the realization of the random variable θ such that $\epsilon_i \sim f_\theta(\epsilon_i), \forall i \in [1, N]$ where f_θ is a stationary probability distribution function.

Now we assume that the realized set of erasure probability $\{\epsilon_1^N\}$ is exposed to neither the encoder nor the decoder. In this case, the only available information that can be extracted to the encoder and the decoder is the set of moments from the given distribution f_θ . We prove that under this condition, Polar codes can achieve the symmetric capacity.

Theorem 2. *For a set of BECs $\{W_{(i)}\}$ with the unknown set of erasure probabilities $\{\epsilon_i\}, i \in [1, N]$, but the distribution f_θ is given to the transceiver, there exist polar codes, for arbitrary small $\delta \leq 0$, that achieve the symmetric capacity I_s, in the sense that as $N \to \infty$, the fraction of indices i satisfies:*

$$\lim_{N\to\infty} \frac{|\{i|I(W_N^{(i)}) \in (1-\delta, 1]\}|}{N} = I_s$$
$$\lim_{N\to\infty} \frac{|\{i|I(W_N^{(i)}) \in [0, \delta)\}|}{N} = 1 - I_s$$

where the symmetric capacity I_s is defined as an average of individual transmit channel's capacities: $I_s(W_N) = \frac{1}{N}\sum_{i=1}^{N} I(W_{(i)})$ where $W_N : X_1^N \mapsto Y_1^N$.

3.1.1 Proof of Theorem 2

By the law of large number, the empirical channel behavior for a codeword would be described by the the first moment ϵ_m of the distribution f_θ. And, as mentioned, since the

transceiver is oblivious to the exact set of erasure probabilities of underlying parallel channels, it has no choice but exploiting ϵ_m into consideration for constructing codewords.

Now, let us consider the meaning of I_s:

$$I_s(W_N) = \frac{1}{N}\sum_{i=1}^{N} I(W_{(i)}) \tag{3.1}$$

$$= \frac{1}{N}\sum_{\imath=1}^{N}(1-\epsilon_i) \tag{3.2}$$

$$= 1 - \frac{1}{N}\sum_{\imath=1}^{N}\epsilon_i \tag{3.3}$$

$$= (1-\epsilon') \tag{3.4}$$

where $\epsilon' = \frac{1}{N}\sum_{\imath=1}^{N}\epsilon_i$.

Hence, we can conclude that the non-identically distributed N parallel BECs model is equivalent to the N identically distributed BECs model whose erasure probability is ϵ' for a large N and we can achieve the symmetric capacity.

Next, it has to be verified whether (3.4) is reliably achievable based on information the transceiver has given. To that end, we have to show following two issues are true:

- The existence of Polar codes over BEC with erasure probability ϵ_m
- $\epsilon' \geq \epsilon_m$ for all distributions f_θ

The first statement is proved to be true in Theorem 2 of [1] in that for any B-DMC W, there exist a sequence of information bit sets A_N such that $|A_N| \geq NR$ with arbitrary small

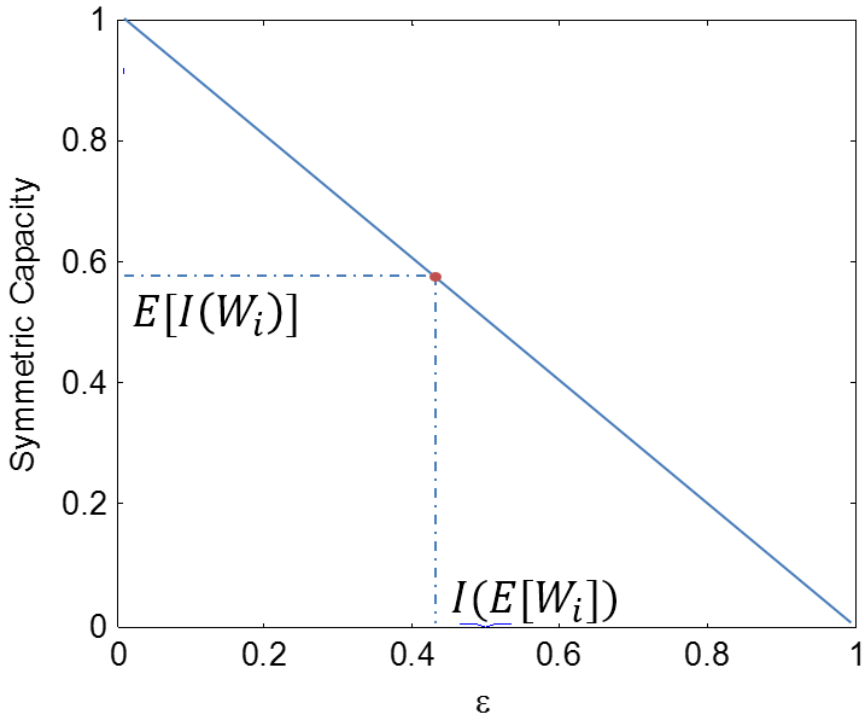

Figure 3.1 The symmetric capacity I_s of BEC

error probability. In this scenario, $|A_N| = \lfloor N(1-\epsilon')\rfloor$.

The second statement is proved through the linearity of I_s to the erasure probability. in Fig. 3.1, I_s is depicted and note that the I_s is an affine function of the erasure probability ϵ. Therefore, $\epsilon' = \epsilon_m$ for any distribution f_θ since $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \epsilon_i = \epsilon_m$.

Note that under erasure channels, the equality in the Jensen's inequality holds:

$$I(E[W_{(i)}]) \leq E[I(W_{(i)})] \tag{3.5}$$

This completes the proof of the Theorem 2.

3.1.2 The Achievable Polar coding scheme

According to the Theorem 2, the achievable Polar coding scheme is straightforward. First, given the distribution f_θ, the encoder calculates the first moment $\mathbb{E}[\epsilon] = \epsilon_m$. Then it construct the message vector u_1^N by figuring out the information index set A_N and with the pre-defined frozen bits u_F. This message sequence is encoded through the generator matrix G_N and is transmitted through the non-identically distributed parallel BECs of $\{\epsilon_1^N\}$. The procedure is summarized in the Algorithm 3.

Algorithm 3 Polar Coding Scheme

Encoding process

1: Given f_θ, calculate ϵ_m
2: Figure out the information sets A_N according to BEC(ϵ_m)
3: Encoding: $x_1^N = u_1^N \cdot G_N$

Decoding process

1: Given (A, u_F, ϵ_m) Perform SC decoding: $y_1^N \rightarrow \hat{u}_A$.

We should note that to achieve the symmetrical capacity under the non identically distributed parallel BECs, with unknown channel parameters, the only constraints that is required is the code length $N \rightarrow \infty$.

3.2 Random Erasure probabilities with non-identical distributions

In this part, we consider the case of N non-identically distributed BECs $W_{(i)}$: for $\forall i \in [1, N]$ that each distribution f_{θ_i} of the erasure probability ϵ_i for each transmit channel $W_{(i)}$ could be different for different bit channel indices. Thus this scenario includes previous scenarios

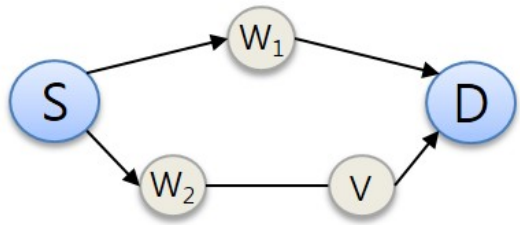

Figure 3.2 Example: the multihop transmission.

as a special case.

As an example, consider multi-hop communications between two nodes in Fig. 3.2. In this figure, each data element in S would be delivered to D through different paths in a multi-hop fashion. If the number of hops are increased, it gets more difficult for S and D to track the exact parameters that models each path. Furthermore, these paths could have statistical variation due to unstable media or interference from different noise sources such as the node V in the figure. In those cases, the proposed scenario makes sense. One simple coding scheme that is able to achieve the symmetric capacity is to convey multiple codewords as a group for each decoding stage. In each stage, we exploit the set of parallel non-identical channels $L = 2^l$ times. The reason of this power of two format is to match with the length of Polar codewords.

Let $W_{(i)}(j)$ mean the j^{th} access to the channel $W_{(i)}$, $\epsilon_i(j)$ as the instantaneous erasure probability of the channel $W_{(i)}(j)$ that follows the distribution f_{θ_i}, and $I_{s,i}(W_{(i)})$ is the symmetric capacity of $W_{(i)}$ by accessing it L times.

In Table 3.1, the instantaneous capacity for L blocks are depicted. The distributions of

erasure probabilities may be different with each other, and each transmit channel (row) has an ergodic behavior as shown in the last column of the table. Here, $\bar{I}(W_{(N)}) = \frac{1}{L}\sum_{j=1}^{L} I(W_{(i)}(j))$.

Table 3.1 Ergodic behaviors of Instantaneous Capacities

	1	2	$\cdots$	L	Avg.
1	$I(W_{(1)}(1))$	$I(W_{(1)}(2))$	$\cdots$	$I(W_{(1)}(L))$	$\bar{I}(W_{(1)})$
2	$I(W_{(2)}(1))$	$I(W_{(2)}(2))$	$\cdots$	$I(W_{(2)}(L))$	$\bar{I}(W_{(2)})$
i	$I(W_{(i)}(1))$	$I(W_{(i)}(2))$	$\cdots$	$I(W_{(i)}(L))$	$\bar{I}(W_{(i)})$
N	$I(W_{(N)}(1))$	$I(W_{(N)}(2))$	$\cdots$	$I(W_{(N)}(L))$	$\bar{I}(W_{(N)})$

The symmetric capacity in this case becomes

$$I_s(W_{[N]}^L) = \frac{1}{L}\sum_{j=1}^{L}\frac{1}{N}\sum_{i=1}^{N} I(W_{(i)}(j)) \tag{3.6}$$

$$= \frac{1}{L}\sum_{j=1}^{L}\frac{1}{N}\sum_{i=1}^{N}(1-\epsilon_i(j)) \tag{3.7}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(1-\frac{1}{L}\sum_{\mathrm{J}=1}^{L}\epsilon_i(j)\right) \tag{3.8}$$

$$\rightarrow \frac{1}{N}\sum_{i=1}^{N}(1-\epsilon_{m_i}) \tag{3.9}$$

where ϵ_{m_i} is the first moment of $\epsilon_i \sim f_{\theta_i}$. Note that $\lim_{L\to\infty} I_s(W_{[N]}^L)\frac{1}{N} = \sum_{i=1}^{N}(1-\epsilon_{m_i})$. The equality is due to the affinity of the symmetric capacity over the domain of erasure probability.

Now let us consider two cases. In the first case, we assume that the encoder is able to be adapted to various code lengths. This means that the encoder can construct generator

matrices G_N for any exponent n ($N = 2^n$). In the second case, the coding structure is fixed, thus parameter N (and the following G_N) can not be changed.

3.2.1 Case1: Variable coding structure

In this case, assume the encoder can exploit any exponent l, and construct $L \times L$ generator matrix G_L where $L = 2^l$. Then the following proposition is satisfied:

Proposition 4. *For a set of non-identically distributed BECs $\{W_{(i)}\}$, with a set of random erasure parameter $\{\epsilon_i\}, i \in [1, N]$ that each ϵ_i follows non identical f_{θ_i}, the symmetric capacity $I_s(W_{[N]}^L)$ is achievable by exploiting multiple streams of Polar codewords.*

To prove the Proposition 4 is the same as to prove the existence of Polar codes that achieve the set of individual capacities $\{\bar{I}(W_{(i)})\}$, since the symmetric capacity $I_s(W_{[N]}^L)$ is the sum of them.

And in Theorem 2, we have proved that there exist polar codes that achieve each capacity $\{\bar{I}(W_{(i)})\}$, in the sense that as $L \to \infty$ through power of two, the fraction of indices $j \in [1, L]$ of the i^{th} message block satisfies:

$$\lim_{L\to\infty} \frac{|\{j|I(W_L^{(j)}) \in (1-\delta, 1]\}|}{L} = \bar{I}(W_{(i)})$$
$$\lim_{L\to\infty} \frac{|\{j|I(W_L^{(j)}) \in [0, \delta)\}|}{L} = 1 - \bar{I}(W_{(i)})$$

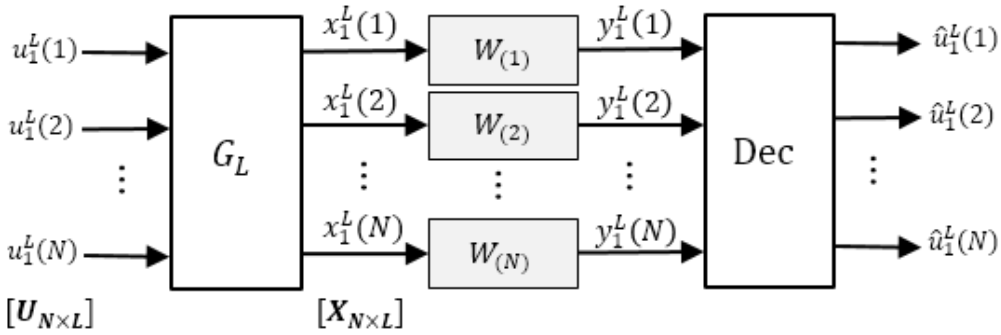

Figure 3.3 Multiple streams of Polar coding structure

for an arbitrary small $\delta \geq 0$, and for all channel index $i \in [1, N]$. Therefore, the Proposition 4 is true for any set of distributions $\{f_{\theta_i}\}$.

3.2.1.1 Polar coding scheme

In Fig. 3.3, the encoding and the decoding procedures are depicted for all transmit channel index. Given all distributions $\{f_{\theta_i} | \forall i \in [1, N]\}$, the encoder calculates the set of first moments $\{\epsilon_{m_i}\}$, and evolved bit-channel capacities $\{I(W_N^{(i)})\}$ with $\{\epsilon_{m_i}\}$ according to those recursive equations (2.19) and (2.20).

With these moments and symmetric capacities, it then constructs N streams Polar codewords $\{x_1^L(i)\}$ of length $L = 2^l$.

To that end, first the encoder defines information set $A_L(i) = \{j | I(W_N^{(j)}) \geq I(W_N^{(k)}), \forall j \in A_L, k \in A_L^c\}$, and whose size is $|A_L(i)| = \lfloor L(1 - \epsilon_{m_i}) \rfloor$

With fixed frozen bits u_F where $F = [1, L] \setminus A_L(i)$, message vectors $\{u_1^L(i)\}$ are stacked in the $N \times L$ message matrix U . It is then encoded with the $L \times L$ generator matrix G_L

and output codewords matrix X . Each column of X would propagate sequentially through the non-identically distributed parallel BECs.

At the receiver, it saves L output vectors in matrix Y and produces estimates applying the SC decoder row by row. This procedures are summarized in Algorithm 4.

Note that there are no constraints on N. Actually, since the proposed coding scheme does not affected by N, it is robust to the deletion of some transmit channels. That is, if some set of channels W_J where J is a subset of $[1, N]$ are lost in that the corresponding symmetric capacities are all zero, the encoder and the decoder will simply decrease N to $N - |J|$, and transmit through $W_{[1,N]\setminus J}$. Then still the symmetric capacity $I_s(W^L_{[N]\setminus J}) = \sum_{\forall i \in [1,N]\setminus J} I(W_{(i)})$ is achievable.

Algorithm 4 Encoding and Decoding Process

Encoding process: Repeat $\forall i \in [1, N]$

1: Calculate $\bar{I}(W_{(i)})$
2: Find ϵ_{m_i} s.t. $I(\epsilon_{m_i}) = \bar{I}(W_{(i)})$
3: Define index set A_i based on (L, ϵ_{m_i})
4: $x_1^L(i) = u_1^L(i) \cdot G_L$
5: Store $x_1^L(i)$ in the i^{th} row of X
6: Transmit each column of X

Decoding Process:

1: Stack N blocks into matrix Y row by row.
2: Given (A_i, L, u_F), perform SC decoder: $Y(i) \mapsto \hat{u}_{A_i}$ $\forall i \in [1, N]$

3.2.1.2 Proof of achievability

From the set of BECs $\{W_{(i)}\}$ for $\forall i \in [1, N]$, the size of each information set$|A_L(i)| = \lfloor L(1 - \epsilon_{m_i}) \rfloor$, the unit of which is bits per L channel uses. It is known that the SC decoder

would recover each message u_{A_i} with vanishing probability of error as $L \to \infty$ for $\forall i$. Now, define an individual rate R_i as $\frac{|A_i|}{L}$, then it converges to $\bar{I}(W_{(i)})$ under the same condition. That is, for any $\delta_i \in [0, 1)$:

$$\begin{aligned} R_i &= \frac{1}{L}\lfloor L(1-\epsilon_{m_i})\rfloor \\ &= (1-\epsilon_{m_i}) - \frac{\delta_i}{L} \\ &\xrightarrow{L\to\infty} \bar{I}(W_{(i)}). \end{aligned}$$

For the symmetric capacity $I_s(W_{[N]}^L)$ is the same as the arithmetical mean of its parts $\{\bar{I}(W_{(i)})|\forall i \in [1, N]\}$: $I_s(W_{[N]}^L) = \frac{1}{N}\sum_{i=1}^{N} \bar{I}(W_{\Lambda_i})$, we can conclude that it is achievable from the proposed scheme. The complexity of this polar coding scheme is $O(NL\log L)$, since it is a concatenation of N SC decoders of length L.

In addition, consider a transmitter that the encoding structure is not able to be changed. Then, we have no choice but utilize the fixed size of $N \times N$ generator matrix G_N where $N = 2^n$ and produced codewords are of the same length N. We can handle this problem by setting L identical to N. The encoder defines the collection of information index sets $\{A_i\}$ from $\{\epsilon_{m_i}\}$, where $|A_i| = \lfloor N \cdot I(\epsilon_{m_i})\rfloor$. Using a common generator matrix G_N, the encoder sequentially produces N of polar codewords $x_1^N(i) = u_1^N(i) \cdot G_N$ for $i \in [1, N]$. These codewords $\{x_1^N(k)\}$ are stacked into as roww of the matrix X. The complexity of this polar coding scheme is $O(N^2 \log N)$, since it is a concatenation of N SC decoders.

3.2.2 Case2: Fixed coding structure

Consider a transmitter that the encoding structure is not able to be changed. Hence, we have no choice but utilize the fixed size of $N \times N$ generator matrix G_N where $N = 2^n$ and produced codewords are of the same length N. As a result L is set to be N.

Theorem 3. *For a set of non-identically distributed BECs $\{W_{(i)}\}$, with a set of random erasure parameter $\{\epsilon_i\}, i \in [1, N]$ that each ϵ_i follows non identical f_{θ_i}, there exist polar codes that achieve a set of average capacities $\{\bar{I}(W_{(i)})\}$, in the sense that as $L \to \infty$ through power of two, the fraction of indices $j \in [1, L]$ of the i^{th} message block satisfies:*

$$\frac{|\{j|I(W_L^{(j)}) \in (1-\delta, 1]\}|}{L} \to \bar{I}(W_{(i)})$$
$$\frac{|\{j|I(W_L^{(j)}) \in [0, \delta)\}|}{L} \to 1 - \bar{I}(W_{(i)})$$

for an arbitrary small $\delta \geq 0$,

We will show that each $\bar{I}(W_{\Lambda_j})$ is achievable by the following coding scheme.

3.2.2.1 Polar coding Scheme

Firstly, given a set $\{f_{\Lambda_j}\}, \forall j \in [1, N]$, the encoder calculates the set of the first moments of capacities $\bar{I}(W_{\Lambda_j})$. It determines the corresponding set of equivalent CPs that describe $\{W'_{\lambda_j}\}$ such that for each W'_{λ_j} satisfies $I(W'_{\lambda_j}) = \bar{I}(W_{\Lambda_j})$. Next the encoder defines the collection of information index sets $\{A_j\}$ from $\{W'_{\lambda_j}\}$, where $|A_j| = \lfloor N \cdot I(W'_{\lambda_j}) \rfloor$. Using a common generator matrix G_N, the encoder sequentially produces M of polar codewords

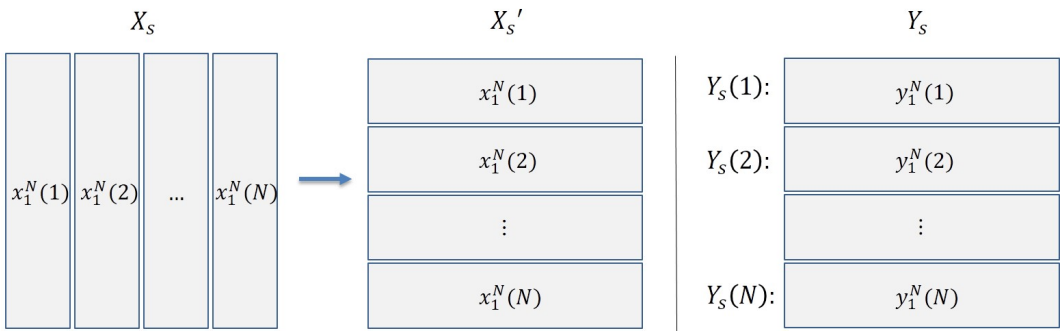

Figure 3.4 The transmitted and received data format.

$x_1^N(k) = u_1^N(k) \cdot G_N$ for $k \in [1, M]$. These codewords $\{x_1^N(k)\}$ are stacked into a matrix X_s. As mentioned before, we set $M = N$ to maintain consistency in length between the output of an interleaver and the number of links. After that, X_s is transposed to X_s', and the columns of X_s' will be transmitted sequentially through parallel channels, which can be performed by an interleaver $\pi(\cdot)$. This procedure is depicted in Fig. 3.4.

The receiver store the N blocks of data into a $N \times N$ matrix Y_s. Then, it divides Y_s into a set of rows, and feed each row vector into the SC decoder element: the j^{th} decoder D_j takes the j^{th} row $Y_s(j)$, and outputs $\hat{u}_{A_j}$. This procedure is summarized in Algorithm 5. The complexity of this polar coding scheme is $O(N^2 \log N)$, since it is a concatenation of N SC decoders.

3.2.2.2 Proof of Achievability

From the set of B-DMCs $\{W_{\lambda_i}'\}$ for $\forall i \in [1, N]$, the size of each information set $|A_j| = \lfloor N \cdot I(W_{\lambda_j}') \rfloor$, the unit of which is bits per N channel uses. It is known that the SC decoder

Algorithm 5 Encoding and Decoding Process

Encoding process

1: Multi-block encoding: M $\leftarrow$ N
2: Calculate $\{\bar{I}(W_{\Lambda_j})\}$ for $\forall j \in [1, N]$,.
3: Find $\{W'_{\lambda_j}\}$ s.t. $I(W'_{\lambda_j}) = \bar{I}(W_{\Lambda_j})$.
4: Define $\{A_j\}$ with $(N, \{W'_{\lambda_j}\})$.
5: Encoding: $x_1^N(j) = u_1^N(j) \cdot G_N$
6: Store $\{x_1^N(j)\}$ in X_s.
7: Interleaver: $X'_s \triangleq \pi(X_s)$
8: Transmit X'_s column by column.

Decoding Process: Given $\{A_j | 1 \leq j \leq N\}$,

1: Stack N blocks into matrix Y_s row by row.
2: For $\forall j \in [1 : N]$, $D_j : Y_s(j) \rightarrow \hat{u}_{A_j}$

would recover each A_j with vanishing probability of error as $N \rightarrow \infty$ for $\forall j$. Now, define an individual rate R_j as $\frac{|A_j|}{N}$, then it converges to $\bar{I}(W_{\Lambda_j})$ as $N \rightarrow \infty$. In terms of sum capacity, for any $\delta \in [0, 1)$, the rate R converges to I_e such that

$$\begin{aligned} R &= \frac{1}{N} \sum_{j=1}^{N} I(W'_{\lambda_j}) - \frac{\delta}{N} \\ &\xrightarrow{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^{N} \bar{I}(W_{\Lambda_j}). \end{aligned}$$

□

Since I_e becomes equivalent to the sample mean of instantaneous capacities when $N \rightarrow \infty$, we can conclude that the sample mean capacity is also achievable by using multiple streams of polar codes.

3.3 Summary

Under the non-independent channel scenario, we assume that N transmit channels are grouped into channels with size r which is a power of two, so that we can deal with the scenario as a non-binary system. If N is not divisible by r ($N \mod r \neq 0$), puncturing may be used to fit the system into a q-ary system. The proposed polar codes appear to be promising for applications where only the knowledge of channel parameter distribution is available, and can be practical for storage applications such as flash memory devices.

Part II

Polar codes schemes for Index Coded Systems

Chapter 4

Nested Polar codes structures for Index codes

4.1 Introduction to Index codes

Channel coding concept is used to mitigate the influence of noise and interferences in the physical layer. In [17], it was also shown that we can get coding gain in higher layers. Compared to the routing and scheduling technique which are devised to prevent bottlenecks of packets from different senders, Alswede et al [18] showed a way of making use of this disadvantage, and showed that the achievable rate can be increased by applying certain in-network processing at an intermediate node when packets are received at the node simultaneously. This type of in-network processing is called network coding. Routing can be treated as a special case of network coding which is a simple permutation. Network coding has received attention since it can enhance system throughput and reliability. For

throughput, network coding technique can take advantages of bottleneck effect of data at the intermediate node in wireless communication to improve the system throughput [19]. Ghaderi et al. [23] has shown that there are reliability benefits by applying network coding technique in their system. Li et. al. [20] showed that the maximum achievable rate can be achieved by linearly combining input packets at an intermediate node. Random linear network coding [21] (RLNC) and opportunistic network coding [22] (OPNC) have been known as one of practical implementations. RLNC randomly chooses elements from a finite field as the coefficients for a linear combination of packets.

OPNC performs bitwise XOR operation of packets that are selected by reception report. RLNC is suitable for the distributed system, and no reception report is needed since it contains all the information in the header to decode the received packets at the receiver node. However, as the number of hops or the number of participants increases, the length of the header also increases, which might degrade the throughput. Although OPNC needs extra report, the portion is not significant compared to the original information, and the implementation of coding and decoding is simple. As a practical implementation of OPNC, Katti et al. [22] introduced a scheme, COPE, that takes advantage of broadcasting nature of wireless communications.

COPE employs practical network coding technique for unicasts in wireless mesh networks to improve total throughput. They showed through experiments that with OPNC in the system, there exist significantly improvements in throughput of wireless networks with UDP

traffic. Recently, Fang et al. [24] gave a analysis of COPE, and argue that the key to COPE's success lies in the interaction between COPE and the MAC protocol. How MAC protocol deals with competing nodes in a given network plays an important role in performance improvement. In this paper, we consider the following two factors. One factor is the channel state information which can affect the performance of a system. The other factor is how to deal with multiple intermediate nodes which can perform network coding simultaneously. This kind of networks, without certain decision methods at the intermediate nodes, we cannot guarantee the throughput gain by using network coding in the system as in [22].

Ming et al. [28] had used uplink model which consists of multiple users, multiple relays and single receiver base station, and had proposed finite field network coding and superposition coding in the relay nodes. In their next paper [29], still with the multiple sender single receiver up-link case, they changed their system model and replaced relay nodes with another user nodes. With this multiple user cooperative communication system, they suggested a diversity network codes scheme over finite fields. Lu et al. [30] used down link model consists of single transmitter base station, single relay node and multiple receiving user nodes. They proposed an instantaneously decodable binary network coding scheme and showed its improved transmission efficiency compared to previously proposed ARQ and network coding based schemes. Bletsas et al. [25] dealt with a cooperative communication system consisting of single source node, single sink node and multiple relay nodes model and introduced a distributed network path selection algorithm which involves opportunistic

relaying to transmit information by using an objective function of channel state at the relay nodes.

The problem of Index Coding with Side Information was introduced by Birk and Kol [34], [35] and it is known any network coding problems can be transformed to equivalent index coding problems [31]. It was motivated by applications such as a satellite communications or large traffic systems as streaming networks. Generally, it consists of a transmitter that broadcasts a set of messages to a set of receiver nodes (or sink nodes).

During the communications between these nodes, each sink node recovers messages with *overhearing* operation: not only the message it requires, but also messages those which are broadcast from the source node. Each receiver saves them in a memory as its side information and reports their information such as message indices to the source.

The transmitter wants to deliver all the set of messages to corresponding sink nodes with a minimum number of transmissions. From the feedback (or Reporting in OPNC) process, every receivers give their pool of information (side information) and its demanded message index. The transmitter exploits this information, and produced coded sequences: the conceptual algorithm is given in [36].

Another possible application of index coding systems is in opportunistic network coding (OPNC) systems, where nodes overhear every data in the wireless channel. This accumulated informations would help to increase the throughput of the system.

Consider the typical Index coding example in Fig. 4.1 with one sender and four receivers.

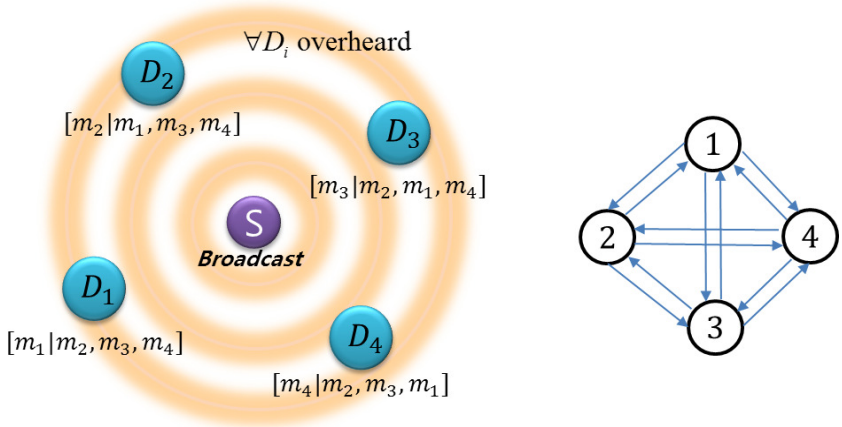

Figure 4.1 This figure depicts a typical scenario that the Index codes have its gain. (a) System model with nodes' demanded message W_i and set of side information K_i : $[W_i|K_i]$. (b) Corresponding directed graph (digraph).

Each receiver D_i demands a independent message m_i and has a set of side information in its memory.

Without Index coding, S has to transmit four times slot for those four messages. However, by utilizing SI, it is known that only one coded transmission is enough. The procedures for calculating the Index codes for this example is as follows:

1. Draw digraph G as Fig. 4.1(b) from SI
2. Represent in a matrix form of $M = A - I$ where A is adjacency matrix that fits G
3. Find minimum rank over $GF(2)$ from M: $minrk_2(G)$ which is the length of index codes
4. Build M' with linearly independent rows of M

According to the 2nd procedure, M is

$$M - I = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \tag{4.1}$$

Since $minrk_2(M) = 1$, the length of index codes is 1, and the only linearly independent row of M is [1,1,1,1]

The resulting index codes c of the example is then

$$C = m_1 \oplus m_2 \oplus m_3 \oplus m_4 \tag{4.2}$$

$$= \oplus_{j=1}^{4} m_j \tag{4.3}$$

Bar-Yossef et al. [36] proved that $\frac{1}{minrk_2(G)}$ is the best rate (thus is defined as the scalar capacity $C_s(G)$) for scalar linear index codes and following theorem holds in general.

For the typical example, $C_s(G) = 1$, since $\frac{1}{minrk_2(G)} = 1$, and the calculated index codes achieves it since the code length is 1.

Theorem 4 ([36]). *For any side information graph G, there exists a linear index code for G whose length equals* $minrk_2(G)$. *This bound is* optimal for all linear index codes for G.

Unfortunately, it has been shown by Peeters [37] that computing $minrk_2(G)$ is $NP - hard$. Several heuristic solutions for this problem were proposed using random coding, composite coding and dynamic programming etc.

4.2 Nested structures for NC and Polar codes

Until now, we reviewed prior works on IC under noiseless channels. However, when the links are noisy, application of channel codes should be considered to overcome erroneous channel effects.

In [38], the authors proposed an error correcting index codes that can correct up to δ [bits], however, it is required large alphabet size for optimal performance. For smaller alphabet, it is still suboptimal.

In [39], a nested coded forward error control coding for binary sequences is proposed. Following their notations, assume that information vectors $\{i_1, i_2, ...i_N\}$ of k bits are jointly encoded to a codeword C of length N with the consideration of SI.

$$C = i_1 G_1 \oplus i_2 G_2 \oplus \cdots \oplus i_N G_N \tag{4.4}$$
$$= \begin{bmatrix} i_1 & i_2 & \cdots & i_N \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_N \end{bmatrix}$$

where each G_k is a generator matrix for i_k such that all of generator matrices are linearly independent. It is called $nested$ due to the encoding structure of (4.4) that includes multiple encoded sequences. The weakness of nested scheme is that search for these orthogonal matrices could be a burden for the system for a large number of sequences of N.

As another candidate channel codes for the index codes, Polar codes has two important features.

1. First, Polar codes is a coset code which means the code structure is much similar to (4.4). Given an index set $A \subset \{1, ..., N\}$ and a binary message vector u_1^N, let $G_N(A)$ be the submatrix of G_N, consist of rows indexed in A. and let u_A be the corresponding subvector of u_1^N. Given such an index set A, and a frozen binary vector u_F of length $N`|A|$, define the polar code $C(N, A, u_F)$ of length N as follows.

   We denote $A^C = F$ the frozen set, and the (fixed thus is exposed to the receiver either) bits u_F frozen bits. The codewords of $C(N, A, u_F)$ are given by

   $$x_1^N = u_A G_A \oplus u_F G_F, \tag{4.5}$$

   and the rate is given by $R = \frac{|A|}{N}$.

2. The nested utilization of Polar codes are known from prior works [40]- [43]

   Define the nested polar code $C(N, A_1, A_2, u_F)$ of length N where $A_2 \subset A_1$ as follows. The codewords of $C(N, A_1, A_2, u_F)$ are the same as the codewords for $C(N, A_1, u_F)$. except for the length. The nested structure is defined by partitioning $C(N, A_1, u_F)$ as cosets of $C(N, A_2, u_{F_2})$, where the entries of u_{F_2} are zero if they correspond to an index in $A_2 \setminus A_1$, and given by the corresponding entry in u_F otherwise. Thus the codewords in $C(N, A_1, A_2, u_F)$ are given by

   $$x_1^N = u_{A_2} G_{A_2} \oplus u_{A_1 \setminus A_2} G_{A_1 \setminus A_2} \oplus u_F G_F \tag{4.6}$$

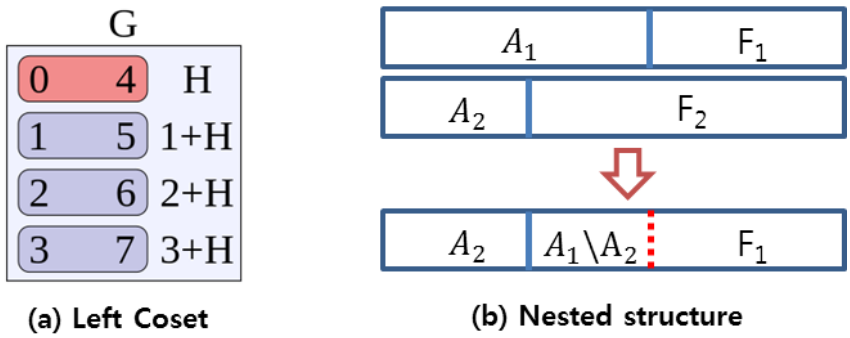

Figure 4.2 (a) Polar codes is one of the coset codes family. This figure depicts an example of the left coset. (b) The nested code structure of Polar codes under degraded channels setting.

where $u_{A_1 \backslash A_2}$ determines which coset the codeword lies in. Note that each coset will be a polar code with A_2^C as the frozen set. The frozen bits u_i are either given by u_F (if $i \in A_1^C$) or they equal the corresponding bits in $u_{A_1 \backslash A_2}$.

The frame structure is depicted in Fig. 4.2(b).

For the simplicity in indices assignment for remaining sections, we categorize channels into to classes: degraded and non degraded classes. When channels from the source S to each sink node D_i are degraded, we denote the order as $W_1 \preceq W_2 \preceq W_3 \preceq \cdots \preceq W_L$, without loss of generality. Under the degraded setting, we can utilize the nested Polar coding scheme if necessary and the inclusion in the nested coding structure is proved from following lemma [44]:

Lemma 4 ([44] Lemma 4.7)**.** *If W_1 is degraded with respect to W_2, then $W_{1,N}^{(i)}$ is degraded with respect to $W_{2,N}^{(i)}$ and $Z_{2,N}^{(i)} \leq Z_{1,N}^{(i)}$*

With the help of this lemma, in next section, we propose a joint ICPC scheme that under the degraded channels environment, the orthogonality constraint is removed.

4.3 ICPC for fully connected SI

4.3.1 General channel setting

In the similar BBC model as Fig. 4.1 where there is one transmitter S and L receivers $(D_1, \cdots D_L)$, and the di-graph of SI is fully connected. Define the rate tuple $R = (R_1, R_2, \cdots, R_L)$ where $\{R_j \leq I(W_j) | \forall j \in [1, N]\}$. The capacity region of BBC is given by the closure of the convex hull of R [45] [46]:

$$\mathbb{C} \triangleq clo\Big(co\big(R(X)\big)\Big) \tag{4.7}$$

over an input distribution $X \sim P_X(x)$. $co(T)$ is a convex hull operation over set T, and $clo(T)$ represents closure of set T. Then the following proposition holds for any BI-DMCs:

Proposition 5. *For any binary input BBC W and symmetric marginal channels $(W_1, W_2, \cdots W_L)$, there exists a Index Coded Polar Coding scheme (ICPC) with fully connected SI that achieves the set of channels' symmetric capacities in* (4.7)*.*

Note that it holds for both degraded and non degraded settings. This proposition is an extended version of [47] where the authors have dealt with two receivers scenario only.

Proof. First, the Index coding solution $\mathcal{C}$ for fully connected SI is calculated

$$\mathcal{C} = \bigoplus_{j=1}^{L} m_j \tag{4.8}$$

since the $minrk_2 = 1$. Each message m_j is inserted into the corresponding FIFO queue q_j.

For a bit channel $W_{j,N}^i$ and a Bhattacharyya parameter $Z_{j,N}^i$ for $j \in [1, L]$ for a marginal channel W_j, we can re-define following information sets:

$$A_{s,N} = \{i | Z_{j,N}^i \leq 2^{-N^\beta}, \forall j \in s \text{ and } Z_{l,N}^i \geq 2^{-N^\beta}, \forall l \in [L] \setminus s\} \tag{4.9}$$

where s is an element of a power set $\mathcal{S}$ without the empty set, $s \subseteq \mathcal{S}$, where $\mathcal{S}$ is defined over the receiver index set $\{1, 2, \cdots L\}$ of size $2^L - 1$, and frozen set $F_{s,N} = A_{s,N}^c$. For example, if $s = \{j\}$ or $\{j, k\}$ then respectively,

$$A_{j,N} = \{i | Z_{j,N}^i \leq 2^{-N^\beta} \text{and } Z_{l,N}^i \geq 2^{-N^\beta}, \forall l \in [L] \setminus j\} \tag{4.10}$$

$$A_{jk,N} = \{i | Z_{j,N}^i \leq 2^{-N^\beta}, \ Z_{k,N}^i \leq 2^{-N^\beta} \text{ and } Z_{l,N}^i \geq 2^{-N^\beta} \forall l \in [L] \setminus \{j, k\}\}.$$

Let the stack top of each queue q_j for the message m_j be $q_j(1)$. However, for the simplicity we denote $q_j(1)$ as q_j since all the other bits in queues are not participate in coding procedures. Then we construct the ICPC $\mathcal{P}(N, A_t, u_{F_t})$ with input bits given by

$$u_i = \begin{cases} \bigoplus_{j \in s} q_j & i \in A_{s,N} \\ u_{F_t} & i \in u_{F_t} \end{cases} \tag{4.11}$$

where $A_t = \bigcup_{\forall s \in S} A_s$ and $F_t = A_t^c$. After each assignment from each queue to the information bit u_i, used queues' stack top should be flushed and updated: $q_j(2) \mapsto q_j(1)$.

At the decoder side, it is known that under the SC decoder, u_{A_t} bits are decoded with arbitrary small error rate as the Polar code length $N \to \infty$. A receiver D_j would recover certain portion q_j of its demanded message m_j by XORing its stored messages in the SI $K_j = [L] \setminus j$; $q_j = d(K_j)$ where $d(\cdot)$ is a decoding function of D_j.

$$\hat{q}_{j,s} = u_{A_j} + \bigoplus_{l \in s} d_s(q_l) \quad \text{for} \quad A_s \tag{4.12}$$

The rate R_j of D_j is

$$R_j = \frac{\sum_{s \in S_j} |A_s|}{N} \tag{4.13}$$

where S_j is all elements of power set S that include index j as their member. Then from the information set structure (4.9), $\sum_{s \in S_j} |A_s| = |A_j|$ for $\forall j \in [1, L]$. Therefore,

$$\begin{aligned} R_j &= \frac{|A_j|}{N} \\ &= \frac{\lfloor N \cdot I(W_j) \rfloor}{N} \\ &= I(W_j) + \frac{\delta}{N} \\ &\xrightarrow{N \to \infty} I(W_j) \end{aligned}$$

Hence we can conclude that the capacity $I_s = (I(W_1), I(W_2), \cdots I(W_L))$ is achievable through ICPC scheme as the Polar code length goes to infinity by achieving individual symmetric capacities.

□

Proposed encoding and decoding procedures are summarized in Algorithm 6 and 7.

Algorithm 6 ICPC Encoding with fully connected SI

procedure SEARCH IC SOLUTION(K_1^L)
    $\mathcal{C} = \bigoplus_{j=1}^{L} m_j$
    Queuing $m_1^L \mapsto q_1^L$
end procedure
procedure ASSIGN INFO INDEX($W_{[L]}, \mathcal{C}$)
    Define $A_{s,N}$
    for $i = 1 : N$ **do**
        if $i \in A_{s,N}$ **then** $u_i = \bigoplus_{j \in s} q_j$
        else $u_i = 0$
        end if
    end for
    Polar encoding: $x_1^N = u_1^N \cdot G_N$
end procedure

Algorithm 7 ICPC Decoding

At node D_j:

procedure SC DEC(N, A_t, u_{F_t})
    $y_1^N \rightarrow \hat{u}_{A_t}$.
end procedure
procedure IC DEC(u_{A_t})
    for $i = 1 : N$ **do**
        if $i \in A_{s,N}$ **then** $\hat{u}_i = \bigoplus_{j \in s} q_j + \bigoplus_{l \in K_j} q_l$
        end if
    end for
end procedure

Example 2. *For L=2, we depicts the indices assignment procedure in Fig. 4.3. Here, indices in 1),2) and 3) subsets are used as the information set A_t. Let $A_{(1)}, A_{(2)}$ and $A_{(3)}$ represent these sets.*

- *For $j \in A_{(1)}$, $u_{A_{(1)}} = q_{1,|A_{(1)}|}$*
- *For $j \in A_{(2)}$, $u_{A_{(2)}} = q_{1,|A_{(2)}|} \oplus q_{2,|A_{(2)}|}$*
- *For $j \in A_{(3)}$, $u_{A_{(3)}} = q_{2,|A_{(3)}|}$*

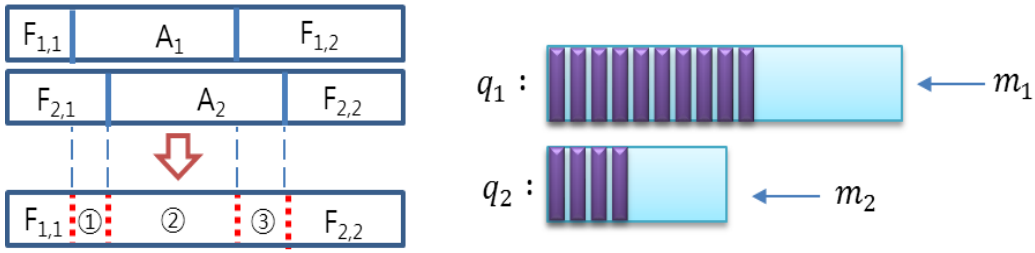

Figure 4.3 Example for 2 receiver under non degraded channel setting. The left figure depicts information indices assignment and the right is FIFO queuing.

Already exploited bits in each queue should be flushed and updated. Decoder D_1 *would recover* $(q_{1,A_{(1)}}, q_{1,A_{(2)}})$ *which is the same as* q_{1,A_1} *and* D_1 *would recover* $(q_{2,A_{(2)}}, q_{2,A_{(3)}})$ *which is the same as* q_{A_2} *as depicted in Fig. 4.3, with arbitrary small error rate.*

Remark 1. *It should be noted that for the correct IC decoding,* D_j *should know all the other's channels* $W_{[L]}$*, or equivalently all the information index set* $A_{[L]}$*. When we denote* χ *as an amount of information that describe a channel* $W_j : S \mapsto D_j$*, the total amount required for the non degraded setting is* χL^2*.*

4.3.2 Degraded channel setting

Contrast to the non-degraded channel case, if channels are degraded in the order of $W_1 \preceq W_2 \preceq W_3 \preceq \cdots \preceq W_L$, the indexing policy becomes much simpler since $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_L$ as depicted in Fig. 4.2. Therefore for $A_{j,N}$ and $A_{jk,N}$ in (4.10), $A_{jk,N} \subseteq A_{j,N}$ for any $j, k \in [1, L]$ and the collection of resulting information sets A_s is

$$A_s = \{A_1, A_2 \setminus A_1, A_3 \setminus A_2, \cdots A_L \setminus A_{L-1}\} \tag{4.14}$$

and we denote $A_{s,j} \triangleq A_j \setminus A_{j-1}$.

The Index coding solution $\mathcal{C}$ is the same $\mathcal{C} = \bigoplus_{j=1}^{L} m_j$ and each message m_j is inserted into the corresponding FIFO queue q_j. Let the stack top of each queue q_j for the message m_j be $q_j(1)$ and for the simplicity we denote $q_j(1)$ as q_j as before. Then we construct the ICPC $\mathcal{C}(N, A_L, u_{F_L})$ with input bits given by

$$u_i = \begin{cases} \bigoplus_{\forall j \in [L] \setminus [j-1]} q_j & i \in A_{s,j} \\ 0 & i \in u_{F_L} \end{cases} \tag{4.15}$$

After each assignment from each queue to the information bit u_i, used queues' stack top should be flushed and updated: $q_j(2) \mapsto q_j(1)$.

At the decoder side, it is known that under the SC decoder, u_{A_t} bits are decoded with arbitrary small error rate as the Polar code length $N \to \infty$. A receiver D_j would recover certain portion of its demanded message m_j by XORing its stored messages in the SI $K_j = [L] \setminus j$.

$$\hat{u}_i = \begin{cases} \bigoplus_{\forall j \in [L] \setminus [j-1]} q_j + \bigoplus_{l \in K_j} q_l & i \in A_s \\ 0 & i \in u_{F_t} \end{cases} \tag{4.16}$$

where '+' operation is still modulo-2 addition.

The rate R_j of D_j is

$$\begin{aligned} R_j &= \frac{\sum_{l=1}^{j} |A_{s,l}|}{N} \\ &= \frac{|A_j|}{N} \end{aligned}$$

As proved in the previous section, the capacity $I_s = (I(W_1), I(W_2), \cdots I(W_L))$ is achievable through ICPC scheme as the Polar code length goes to infinity by achieving individual symmetric capacities.

Remark 2. *For correct IC decoding, D_j should know channels $W_{[j]}$, or equivalently all the information index set $A_{[j]}$. When we denote χ as an amount of information that describe a channel $W_j : S \mapsto D_j$, the total amount required for the non degraded setting is $\frac{\chi L(L+1)}{2}$ which is strictly smaller than the non-degraded case's: $\frac{\chi L(L+1)}{2} \leq \chi L^2$.*

Remark 3. *Larger Index coding gain is achievable compare to the non-degraded setting's due to the inclusions among information index sets.*

4.3.3 IC gain analysis

In this part, we analyze the gain of IC in both degraded and non-degraded settings.

Lemma 5. *Let IC gains be g_d, g_{nd} for degraded and non-degraded settings respectively. Then from* (4.9)*,*

$$g_d = 1 + \frac{\sum_{i=1}^{L-1} k_i}{k_L} \tag{4.17}$$

$$g_{nd} = \frac{\sum_{i=1}^{L} k_i}{\sum_{s \in \mathcal{S}} k_s} \tag{4.18}$$

where L is the number of receivers, $k_i = |A_i|$ of B-DMC W_i and $\mathcal{S}$ is a power set of receiver indices as defined in (4.9)*.*

Proof. The proof of Lemma 5 is easily done from the codeword structure. From the Proposition 5, we proved that there exists ICPC that achieves channel capacities with arbitrary small error rate. It means that each sink D_j would decode u_{A_j} w.p. 1 as $N \to \infty$ where $|A_j| = k_j$. When the SI digraph is fully connected, D_j would recover $q_{|A_j|}$ of length k_j [bits] correctly using its SI K_j.

Hence the total amount of information delivered to receivers is $\sum_{i=1}^{L-1} k_i$ [bits] in both channel settings. In the degraded setting, this amount is transferred through k_L [bits], thus (4.17) holds. And in the non degraded setting, it is assigned into $\sum_{s \in \mathcal{S}} k_s$ [bits] IC codeword, hence (4.18) holds. □

We have following features to notice for these gains

- When channels are noiseless, g_d becomes L. Note that when $k_i = k_j = k$ for $\forall i, j \in [1, L]$, corresponding $g_d = L$ either, which holds when all the channels are identical such that $W_i = W_j = W$. Hence, the reduced gain of $1 - \frac{\sum_{i=1}^{L-1} k_i}{k_L}$ is induced from non-identical channels environment.

- The overall rate R_t, including the index coder gain g above and the Polar coder rate R_p, is then for the degraded case under full SI and for the non degraded case respectively:

$$
\begin{aligned}
R_t &= g_d \cdot R_p \\
&= \frac{\sum_{i=1}^{L} k_i}{\sum_{s\in\mathcal{S}} k_s} \cdot \frac{\sum_{s\in\mathcal{S}} k_s}{N} \\
&= \frac{\sum_{i=1}^{L} k_i}{N} \\
&= \sum_{i=1}^{L} \frac{k_i}{N} \\
&= \sum_{i=1}^{L} I(W_i)
\end{aligned}
\qquad
\begin{aligned}
R_t &= g_{nd} \cdot R_p \\
&= \frac{\sum_{i=1}^{L} k_i}{k_L} \cdot \frac{k_L}{N} \\
&= \frac{\sum_{i=1}^{L} k_i}{N} \\
&= \sum_{i=1}^{L} \frac{k_i}{N} \\
&= \sum_{i=1}^{L} I(W_i)
\end{aligned}
$$

which again represent that *ICPC scheme under fully connected SI achieves sum-capacity of BBC, independent of channel settings.*

Recall that one of the difference is the availability of channel information to the decoders: non-degraded setting requires much more amount of channel information for each receiver.

- For non degraded case, if all channels are identical, then $g_{nd} = L$ either. (not from the equivalence in k).

- $g_{nd} \leq g_d$

  This property can be easily verified since g_{nd} can be re-expressed as $1 + \frac{\sum_{i=1}^{L-1} k_i - \alpha}{k_L + \alpha}$ which is obviously smaller than g_d.

4.4 ICPC for Arbitrary SI

Let the linear IC length be $minrk_2(M) = r$ for some $L \times L$ matrix M that fits digraph of SI where L is the number of receivers, and the corresponding index codes are denoted as $\mathcal{C} = (c_1, c_2, \cdots, c_r)$.

In this section, we consider whether there exist ICPC schemes that achieve the sum rate $R_t = \frac{1}{r}\sum_{i=1}^{L} I(W_i)$ for arbitrary side information patterns, and in what condition there exist with probability 1. Note that such R_t is optimal from two reason. First, for the Linear Index Coding (LIC) theory, the minimum rank of the matrix M is identical to the shortest length of the LIC, which means the source has to transmit codewords at least r times to satisfy demands of L receivers. Second, due to underlying noisy channels, the maximum rate is the symmetric capacity $I(W_i)$ for a B-DMC W_i. Therefore, in the joint ICPC process, the optimal rate would be the joint of these two limits, which is $R_t = \frac{1}{r}\sum_{i=1}^{L} I(W_i)$. Through ICPC schemes deliver information of amount of equal to the $Cut - capacity$ through r transmissions.

Now we investigate whether there exist an ICPC scheme that achieves such R_t for arbitrary L and arbitrary SI patterns. We figured out two important results according to this issue.

- For some SI pattern, there exist feasible IC solutions; not every candidate IC solutions are qualified for the ICPC scheme. There could be at most $\binom{L}{r}$ combinations in

choosing linearly independent rows in M. However, some combinations may not be feasible in a sense that decoding error rates in at least one receiver would not converge to zero, even when $N \to \infty$ under corresponding ICPC transmissions.

For example, for $L = 3, r = 2$ and under the degraded channel structure, assume M as follows:

$$M = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Denote rows of M as $g_i, i \in [1, L]$. The feasible set of rows is g_1, g_2, and corresponding IC words are $\mathcal{C} = (c_1, c_2)$ where $c_1 = 1 + 3$, $c_2 = 1 + 2$. If we choose g_2, g_3 instead, the receiver D_1 would fail to reliable decoding since $m_2 \notin K_1$.

- In general, for arbitrary L and SI patterns, there are counter examples that fail to achieve $R_t = \frac{1}{r} \sum_{i=1}^{L} I(W_i)$ under the non-degraded channel setting, and even in the degraded case. For example, let $(L, r) = (4, 3)$ and assume the degraded structure with M as follows:

$$M = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

The third row is the modulo-2 sum of the first and the second row. Hence, the set of linearly independent row indices are 4 and another two among $\{1, 2, 3\}$. In every

cases, there exists one receiver whose decoding is unreliable. If (1,2) is chosen, D_3 becomes unreliable, for (2,3) D_1 is unreliable and for (1,3) D_2 becomes unreliable in decoding.

In this paper, we discovered constraints to ensure the existence of such achievable ICPC schemes for arbitrary L under the degraded and the non-degraded cases. To say the conclusion first, in degraded channel structure, there exist such ICPC schemes $w.p.1$ for $L = 2, 3$, and from $L \geq 4$, they exist if the constraint for the degraded case is satisfied. And under the non-degraded setting, there exist such ICPC schemes $w.p.1$ for $L = 2$, and from $L \geq 3$, they exist if the constraint for the non degraded is satisfied.

L	$L = 2$	$L = 3$	$L \geq 4$
Deg.	O	O	Cond.
Non-deg	O	Cond.	Cond.

This table summarizes the existence of ICPC schemes that achieve $R_t = \frac{1}{r} \sum_{i=1}^{L} I(W_i)$ in each channel setting. Here Cond. means it exists conditionally following Constraint 1 for Deg. and (2)

- **Constraint: Degraded**

  Define an index set B whose elements are chosen linearly independent row indices of

M consists of L rows $\{g_1, g_2, \cdots, g_L\}$ as index coding encoding vectors.

$$g_j(j:L) = f(g_B) \tag{4.19}$$

$$= \sum_{j=1}^{r} \alpha_j \cdot g_{B(j)} \tag{4.20}$$

$$= \sum_{j \in B'} g_{B'} \tag{4.21}$$

where $\alpha_j \in GF(2)$. We can represent g_j using the set of non-zero α_j in B which is defined as B'.

For each D_j ($j \in B^c$) to be reliably decode ICPC word, the following should be hold (or equivalently for each non-chosen row g_j ($j \in B^c$)) :

For $\forall k \in [1, L-j]$

Condition 1. *For $\forall k \in [1, L-j]$, if $\exists g_{B_j}(j+k) = 1$, then $m_{j+k} \in K_j$. Else $\forall g_{B'}(j+k) = 0$ in B', there are no constraints for g_j, hence D_j would decode correctly w.p.1*

if it is not a repetition of some g_i ($i \in B$). If there is at least one feasible B satisfying the corresponding condition, the existence of R_t achievable ICPC schemes is assured. Note that if a row g_j ($j \in B^c$) is a repetition of $g_i (i \in B)$ such that $g_j = g_i$, then D_j would decode ICPC words reliably. Now we claim that followings are hold under these conditions.

Example 3. *Let us consider following M_1 and M_2.*

$$M_1 = \begin{bmatrix} 1 & b & 1 & 0 \\ a & 1 & 0 & 0 \\ \bar{a} & \bar{b} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad M_2 = \begin{bmatrix} 1 & b & 1 & 1 \\ a & 1 & 0 & 1 \\ \bar{a} & \bar{b} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

In both cases S chooses $B = \{1, 2, 4\}$ as IC encoding vector. In M_1, D_3 would reliably decode ICPC hence given set B is feasible, however in M_2, D_3 would fail in SC-dec since it has no information on q_4 (or m_4) thus B is not a feasible solution. Therefore S has to check another possible B to figure feasible one out. In this sense, if $b = 1$, then feasible $B = \{2, 3, 4\}$ instead which satisfies the above constraint.

- **Constraint: Non-deg.**

  With the same notations as above, each non-chosen row g_j ($j \in B^c$) the following should be hold:

  Condition 2. *For $\forall k \in [1, L]$, if $\exists g_{B_j}(k) = 1$, then $m_k \in K_j$*

In the Condition 1 for degraded and 2 for non-degraded case, we limit both conditions to be satisfied only for non-repetition rows. In some cases of M, non-chosen rows could be identical to one in an set B (repetition). Equivalently, it happens when there are receivers that have identical side information. Therefore, when we transmit an ICPC word designated to a certain subset of receivers, whose encoding vector is originated from a row in B, it naturally

would satisfy demands of receivers those who have identical side information. Hence, we can ignore the decoding availability of repeating rows in B^c.

Following lemma, theorem and corollary are main results of this section.

Lemma 6. *Under the degraded channel setting with $L \leq 3$, there exist a ICPC scheme with probability 1 that achieves $R_t = \frac{1}{r}\sum_{i=1}^{L} I(W_i)$ where $r = minrk_2(M)$.*

Theorem 5. *Under the degraded channel setting with $L \geq 4$, there exist a ICPC scheme that achieves $R_t = \frac{1}{r}\sum_{i=1}^{L} I(W_i)$ if the IC solution is feasible in a sense that non-repeating rows g_j ($\forall j \in B^c$) of M satisfy Condition 1.*

Corollary 2. *Under the non-degraded channel setting with $L \geq 3$, there exist a ICPC scheme that achieves $R_t = \frac{1}{r}\sum_{i=1}^{L} I(W_i)$ if the IC solution is feasible in a sense that non-repeating rows g_j ($\forall j \in B^c$) of M satisfy Condition 2*

Both conditions ensure whether the chosen index coding solution is feasible or not. In other words, we have to choose IC encoding vectors in M that satisfy those constraints. If there is no such a set of rows in M, the achievable rate will be decreased to $\frac{1}{r+\eta}\sum_{i=1}^{L} I(W_i)$ where $\eta \in \mathbb{Z}_+$ denote the additional transmission of words to fulfill all demands of receivers.

Note that if all of each g_j ($\forall j \in B^c$) are repetitions of g_i ($\forall i \in B$) the problem becomes trivial: there always exist a ICPC scheme that achieves $R_t = \frac{1}{r}\sum_{i=1}^{L} I(W_i)$ for all L.

4.4.1 Proof of the Lemma 6

In this part, we verify the existence of ICPC schemes for $L \leq 3$. For the case of numbers of M for L, there are $2^{L(L-1)}$ cases. Thus, it is quite straightforward for $L = 2$ since there are only 4 cases to check. For $L = 3$, though there are 64 cases, we can check ones whose $r = 2$ cases only, since $r = 1, 3$ cases are trivial; 1 represents the Full SI model which the existence and the achievable ICPC scheme is developed already, and $r = 3$ indicates there are no IC gain, hence simple non-coded transmissions could be be performed. However, the rate for $r = 3$ may be improved under the degraded channel setting with ICPC scheme, and it goes same with $(L, r) = (2, 1)$. Let us start from $L = 2$.

Table 4.1: ICPC for $L = 2$, Deg.

M	$\mathcal{C}$	R_{deg}
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$c_1 \mapsto (k_1, \phi)$ $c_2 \mapsto (\phi, k_2)$	$\frac{1}{2N}(k_1 + k_2)$
$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$	$c_1 \mapsto (k_1, \phi)$ $c_2 \mapsto (\phi, k_2)$	$\frac{1}{2N}(k_1 + k_2)$
$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$c_1 \mapsto (k_1, k_2 - k_1)$ $c_2 \mapsto (\phi, k_2)$	$\frac{1}{2N}(2k_2) = \frac{k_2}{N}$ (q_2 update)
$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	$c_1 \mapsto (k_1, k_2)$	$\frac{1}{N}(k_1 + k_2)$

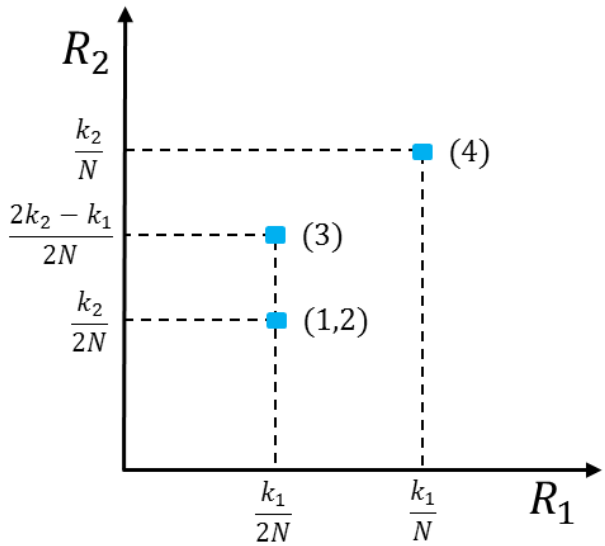

Figure 4.4 Achievable rates using ICPC for $L = 2$ under degraded setting. Note that the point (3) is not allowed in the Non-degraded setting.

In Table 4.1, we list all case of numbers in $L = 2$ for the degraded case. Under degraded channel case: $W_1 \preceq W_2$ and hence $k_1 \leq k_2$. The first three Ms are rank of 2 thus no IC gain is expected, and only the last case's rank is 1 where IC is possible. The R is the rate the system can achieve by using ICPC schemes. One can easily check that there are ICPC schemes that achieve $R_t = \frac{1}{r}(I(W_1) + I(W_2))$ for all possible M. Hence we can conclude that Lemma 6 is true for $L = 2$.

In addition, we observe that with the use of ICPC under the degraded channel structure, we can get additional gain in the rate even when there is no IC gain. This gain can be acquired if we $update$ the second queue q_2 for m_2 after creation of c_1. In Fig. 4.4, we depict these rates.

Table 4.2: ICPC for $L = 2$, Non-deg

M	$\mathcal{C}$	R_{non}
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$c_1 \mapsto (k_1, \phi)$ $c_2 \mapsto (\phi, k_2)$	$\frac{1}{2N}(k_1 + k_2)$
$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$	$c_1 \mapsto (k_1, \phi)$ $c_2 \mapsto (\phi, k_2)$	$\frac{1}{2N}(k_1 + k_2)$
$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$c_1 \mapsto (k_1, \phi)$ $c_2 \mapsto (\phi, k_2)$	$\frac{1}{2N}(k_1 + k_2)$
$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	$c_1 \mapsto (k_1, k_2)$	$\frac{1}{N}(k_1 + k_2)$

In Non-degraded case, IC gain exist only when Full SI is guaranteed since there is no inclusion among information sets as in the degraded case.

For $L = 3$, we check for both degraded and non-degraded cases. There are 64 cases of SI patterns that consist of 1 full SI pattern, 29 of r=2 and 34 of r=3 patterns. As mentioned patterns of $r = 1$ and $r = 3$ are trivial: the achievable ICPC scheme for full SI is proposed in previous section, and for those who have no IC gain, it is merely a non-coded transmissions. Though, as we observed in $(L, r) = (2, 2)$ of Deg. case, there could exist extra gains in some patterns among $(L, r) = (3, 3)$ due to degraded channel structure, we do not consider those

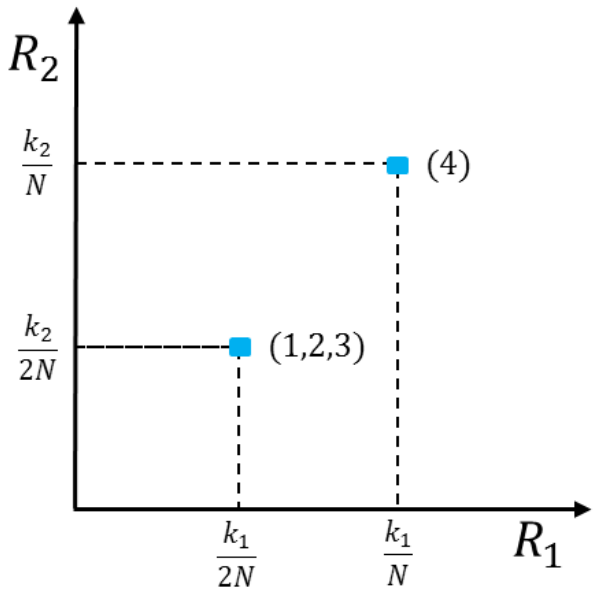

Figure 4.5 Achievable rates using ICPC for $L = 2$ under Non-deg. setting. The gain exist only for the full SI case.

minor effect.

Hence we should verify 29 patterns of $r = 2$. Those patterns is categorized into three cases according to the choice of linearly independent rows in M.

$$\begin{bmatrix} g_1 \\ g_2 \\ * \end{bmatrix} \quad \begin{bmatrix} g_1 \\ * \\ g_3 \end{bmatrix} \quad \begin{bmatrix} * \\ g_2 \\ g_3 \end{bmatrix} \tag{4.22}$$

Start from the degraded channel structure. In the leftmost case in (4.22), note that as W_3 is the best channel, it would reliably perform the SC-decoding. Since g_3 is a linear combination of g_1 and g_2, IC decoding also be successful w.p. 1, once SC-dec result is correct. Therefore,

$\exists$ ICPC scheme, achieves R_t for this category. For those two categories, we can easily verify the achievable ICPC schemes by considering all three linear combinations of each case. Hence, for $L = 3$ degraded case, $\exists$ ICPC scheme, achieves R_t, which completes the proof of Lemma 6.

For non-deg. case, there are two counter examples out of 64 patterns that fail to achieve R_t via ICPC scheme.

4.4.2 Proof of the Theorem 5

The constraint is equivalent to that D_j for $j \in B$ has messages from m_{j+1} to m_L as its side information. Let the minimum rank of M as $minrk_2(M) = r$.

- For $\forall D_j (j \in B)$, there is a ICPC word consists of information that D_j has as SI. Hence as a result, it is given $P(N, A_j, u_{F_j})$ from the degraded nature, so that reliable SC-dec decoding is guaranteed as $N \to \infty$.

- Assume for $\forall D_j (j \in B^c)$, Constraint 1 for Deg. is satisfied. To guarantee the reliable SC-decoding, each should be given $P(N, A_j, u_{F_j})$ and under the degraded channel structure,

$$A_{j+l} \setminus A_{j+l-1} \subset F_j$$

for $\forall l \in [1, L - j]$.

First, since we assumed that $g_j(j : L) = \mathbf{1}_{L-j+1}$ which is equivalent to $\{m_{j+1}, \cdots, m_L\} \subseteq$

K_j, D_j knows all u_T where $T = A_{j+l} \setminus A_{j+l-1}$ for all l. Therefore, D_j can perform SC-decoding reliably, $P_e \to 0$, given $P(N, A_j, u_{F_j})$ as $N \to \infty$.

Second, now we have to verify that for $g_j = f(g_B)$ where f is a linear function, it is uniquely reversible w.p.1 which ensures reliable IC decoding, after the successful SC-decoding procedure.

Recall that each message sequence m_j is inserted into the corresponding FIFO queue q_j of length $k_j = |A_j|$. Each queue is divided into consecutive subsets: q_j consists of j-subsets from q_{j1} to q_{jj} whose lengths are $k_1, (k_2 - k_1), \cdots, (k_j - k_{j-1})$ respectively.

From M, S chooses feasible r linearly independent rows, satisfying Constraint 1 for Deg., and denote them as $\alpha_\eta = [\alpha_{\eta 1}, \alpha_{\eta 2}, \cdots, \alpha_{\eta L}]$ where $\alpha_{\eta j} \in GF(2)$ for $\forall \eta \in [1, r]$ and $\forall j \in [1, L]$.

Then the η^{th} IC word c_η in $\mathcal{C}$ which would be mapped to information bits of the η^{th} Polar codeword u_{A_η} also consists of serial concatenated sub-vectors from $c_{\eta 1}$ to $c_{\eta L}$ whose length are $k_1, (k_2 - k_1), \cdots, (k_j - k_{j-1})$ respectively.

Each $c_{\eta j}$ is then calculated as follows:

$$c_{\eta j} = \sum_{\nu=j}^{L} \alpha_{\eta j} \cdot q_{\nu \eta} \tag{4.23}$$

where the summation is mod-2 addition. This IC word construction for the ICPC scheme is depicted in Fig. 4.6.

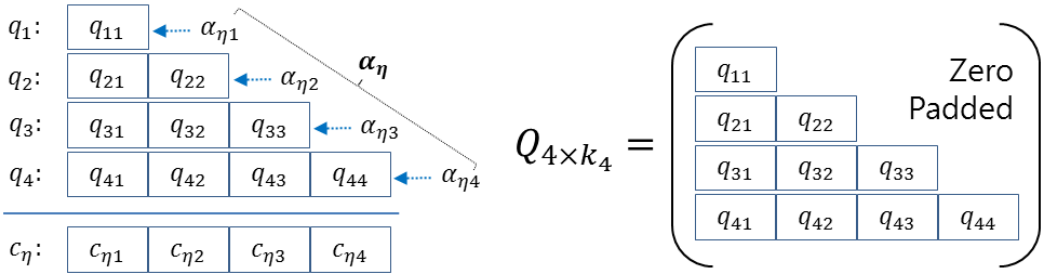

Figure 4.6 The IC word construction of $L = 4$ that is modified for ICPC scheme and its matrix format Q. The size of Q is $L \times k_L$.

Each D_j should recover q_j by exploiting SI K_j and overheard data. We will show that there exists a vector $\beta_j = [\beta_{j1}, \beta_{j2}, \cdots, \beta_{jr}]$ for each D_j such that

$$\beta_j \diamond \mathcal{C} = g_j * Q \tag{4.24}$$

where matrix Q is defined from queues and IC words $\mathcal{C} = [c_1, c_2, \cdots, c_r]$. The $\diamond$ operation is similar to the inner product with slight abuse of notation as follows:

$$\beta_j \diamond \mathcal{C} = \sum_{\eta=1}^{r} \beta_{j\eta} \cdot c_\eta \tag{4.25}$$

$$g_j * Q \triangleq \begin{bmatrix} g_{j1}q_{11} & 0 & 0 & 0 \\ g_{j2}q_{21} & g_{j2}q_{22} & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ g_{jL}q_{L1} & g_{jL}q_{L2} & \cdots & g_{jL}q_{LL} \end{bmatrix} \tag{4.26}$$

or equivalently each row of $g_j * Q$ can be represented as $[g_{j1}q_1; g_{j2}q_2; \cdots; g_{jL}q_L]$.

By expanding (4.25)

$$< \beta_j \cdot \mathcal{C} > = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_r \end{bmatrix} \tag{4.27}$$

$$= \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1L} \\ c_{21} & c_{22} & \cdots & c_{2L} \\ \vdots & \vdots & \vdots & \vdots \\ c_{r1} & c_{r2} & \cdots & c_{rL} \end{bmatrix} \tag{4.28}$$

$$= \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1L} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2L} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{r1} & \alpha_{r2} & \cdots & \alpha_{rL} \end{bmatrix} \begin{bmatrix} \\ Q \\ \\ \end{bmatrix} \tag{4.29}$$

$$= \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_r \end{bmatrix} * Q \tag{4.30}$$

Recall that each α_i is one of rows in $\{g_1^L\}$ and each non-chosen row g_i is a linear combination of chosen rows: $g_i = f(g_B)$. Since $g_B \mapsto \{\alpha_1^r\}$, we can conclude that

there exists $(\beta_j)_1^r$ w.p. 1. such that

$$g_j = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_r \end{bmatrix} \tag{4.31}$$

The α matrix is acquired from received and overheard $\mathcal{C}$.

Note that once $(\beta_j)_1^r$ exist for each D_j, the remaining IC decoding is trivial since the resulting combination (vector) is the linear combination of all data in its own SI storage with its desired data vector q_j.

From these two statement, we can conclude that each receiver $D_j (j \in [1, L])$ would recover $k_j[bits]$ reliably through r-ICPC transmissions under the Constraint 1 for Deg. and thus achieves

$$\begin{aligned} R_t &= \frac{k_1 + k_2 + \cdots + k_L}{rN} \\ &= \frac{1}{r} \sum_{i=1}^{L} \frac{k_i}{N} \\ &\xrightarrow{N \to \infty} \frac{1}{r} \sum_{i=1}^{L} I(W_i) \end{aligned}$$

in sum sense.

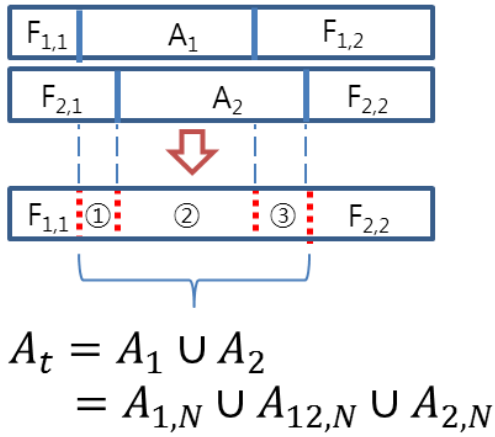

Figure 4.7 Information set A_t of a Polar codeword for some IC solution c_j can be expressed into two ways.

Example 4. *For $(L, r) = (5, 3)$, degraded channels, let M be*

$$M = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix} \tag{4.32}$$

We can see that $g_3(3:5) = \forall 1$, and from the degraded feature, we know that D_5 would reliably perform SC-decoding w.p. 1. Therefore, $B = 1, 2, 4$ is chosen as encoding vectors for the IC procedure.

4.4.3 Achievable ICPC scheme for degraded structures

4.4.3.1 ICPC Encoder

We verify the achievability by suggesting achievable ICPC scheme.

Given a digraph from SI K_j for $\forall j \in [1, L]$, S builds a matrix M that fits the digraph and constructs IC solution $\mathcal{C} = (c_1, c_2, \cdots, c_r)$ of $r = minrk_2(M)$. Each message m_j is inserted into the corresponding FIFO queue q_j.

The IC solution c_j is the XOR sum of some subset of messages whose index set is denoted as J. Denote the power set of J as $S(J)$, and an index set $A_{s,N}$ ($s \in S(J)$) as defined in (4.9).

The indices of Polar codeword for each c_j consist of the information set A_t and the frozen set F_t where $A_t = \cup_{i \in J} A_i$ such that A_i is the information index set of $W_i : S \mapsto D_i$. In non-degraded channel setting, $A_t = \cup_{s \in S} A_{s,N}$, depicted in Fig. 4.7.

Now we assign data bits in queues into A_t and construct the ICPC $\mathcal{C}(N, A_t, u_{F_t})$ in the same assignment policy in (4.11):

$$u_{A_{s,N}} = \bigoplus_{j \in s} q_j(k_{s,N}) \tag{4.33}$$

$$u_{F_t} = 0 \tag{4.34}$$

where $q_j(k_{s,N})$ is length $k_{s,N} = |A_{s,N}|$ sub-vector from stack top. Contrast to (4.11) for full SI, in arbitrary SI case, bits in queues should not be flushed until all transmission is over.

4.4.3.2 ICPC Decoder

Decoding operation also similar to (4.12) except for two points

- No flushing in queues until the end of transmission of $\mathcal{C}$.
- Each receiver *overhears and decodes* all raw q and stores them into its SI storage K'_j, those which are required for its IC decoding. K'_j temporarily stores decoded (from overhearing) q-vectors, by using K_j.

It is known that under the SC decoder, u_{A_t} bits are decoded with arbitrary small error rate as the Polar code length $N \to \infty$. A receiver D_j would recover certain portion of its demanded message m_j part by part via XORing its stored messages in the SI K_j and K'_j. The IC decoder at D_j recovers $q_j(k_j)$ after receiving sufficient size of K'_j for its decoding as follows:

$$d_j(\mathcal{C}) : u_{A_t} \mapsto q_j(k_j) \tag{4.35}$$

Note that once there are no errors from the SC-decoder, the IC decoder d_j would recover $q_j(k_j)$ with probability 1. Since the IC solutions $\mathcal{C} = (c_1, c_2, \cdots, c_L)$ is based on the basis row vectors of M, u_{A_t} of each c_j consist of linear combinations of sub-vectors of those basis. Therefore once D_j gets $\mathcal{C}$ correctly (equivalently, once $N \to \infty$), $q_j(k_j)$ is always to be recovered using local encoding vectors in headers for $\mathcal{C}$.

4.4.3.3 Achievable rate

At the end of the ICPC transmission, all the receivers would recover $(k_1, k_2, \cdots, k_L)$. Hence we the rate R_t is

$$\begin{aligned} R_t &= \frac{\sum_{i=1}^{L} k_i}{rN} \\ &= \frac{1}{r} \sum_{i=1}^{L} \frac{k_i}{N} \\ &\xrightarrow{N \to \infty} \frac{1}{r} \sum_{i=1}^{L} I(W_i) \end{aligned}$$

4.4.4 Proof of the Corollary 2

The proof of the Corollary 2 directly follows the proof of the Theorem 5. The difference is that in the Non-degraded structure, SC-dec of D_j where $j \in B^c$ would successfully decode only when D_j has all messages combined in $\mathcal{C}$.

Note that when B-DMCs are identical, $W_i = W_j$ for $\forall i, j \in [1, L]$, then their information sets for the Polar encoder are identical either; $A_i = A_j$ for $\forall i, j \in [1, L]$.

In this case, the ICPC scheme is merely independent usages of Index codes and Polar codes: since set differences of any pair among $\{A_1 \cdots A_L\}$ are all empty sets, index coded words are simply parsed in the length of $k = |A|$ and would be loaded to the message vector u_A of the length N Polar codeword, without any modification. Therefore the achievable rate R_t of the ICPC scheme would converge to $\frac{kL}{rN}$ as $N \to \infty$.

Condition 1 represents that D_j ($j \in B^c$) should have q_{j+k} as its side information if it is

combined in the u_{F_j} for reliable ICPC decoding.

For each $D_j(j \in B)$, there is one ICPC word which is generated based on K_j which means it knows exact u_{F_j} by exploiting K_j. Therefore, a reliable decoding of u_{A_j} (and q_j) is guaranteed as $N \to \infty$ given a parameter set (N, A_j, u_{F_j}). For reliable decoding of ICPC words in $\forall D_j(j \in B^c)$, we should prove the reliability of both d_S and d_I, consisting d_j under Condition 1. To guarantee the reliable SC-decoding d_S, (N, A_j, u_{F_j}) should be given to d_j, and for the reliable index decoding d_I, the existence of a unique operation should be proved that would reverse the linear combination used in constructing $U_A = \{u_{A_{\{\eta\}}} | \forall \eta \in [r]\}$ which is a set of information bits for PC.

Note that it is necessary for D_j to decode part of U_A since g_{B_k} create g_k (not g_B). We denote them as $U_{A,j} = \{u_{A_{\{i\}}} |$ for some $i \in [r]\}$ and corresponding encoding vectors those that create each elements in $U_{A,j}$ in (4.37) as $\alpha_{\{j\}}$. For example, if D_j requires only $u_{A_{\{2\}}}$ and $u_{A_{\{4\}}}$ in decoding q_j then $\alpha_{\{j\}} = \{\alpha_2, \alpha_4\}$ Let the l^{th} element of $\alpha_{\{j\}}$ as $\alpha_{\{j,l\}}$. If $\alpha_{\{j,l\}}(k) = 1$ for some $k \in [L]$, it means q_k is combined in $U_j(l)$. Since q_k could be added not only to $U_{A,j}(l)$ but also to frozen bits $U_{A^c,j}(l)$, to remove the effect, m_k should be in K_j, which is what Condition 2 means. Under the degraded structure, due to inclusions in information index sets $A_i \subseteq A_{i+1}$ for $\forall i \in [L-1]$ [44], the constraint becomes simpler. If $\alpha_{\{j,l\}}(j+k) = 1$ for some $k \in [L-j]$, K_j should have m_{j+k}, as Condition 1 stands for. This completes the first proof for d_S.

Next, we prove the existence of a vector $\beta_j \in GF(2)^r$ for each D_j such that

$$\beta_j \diamond U = g_j * Q \tag{4.36}$$

where $\beta_j \diamond U = \sum_{i=1}^{r} \beta_{ji} \cdot u^N(i)$. From (4.37), we can get

$$\begin{aligned}
\beta_j \diamond U &= \sum_{i=1}^{r} \beta_{ji} \cdot (\alpha_i * Q) \\
&= \beta_j \diamond (M_B \cdot Q) \\
&= (\beta_j \diamond M_B) * Q
\end{aligned}$$

where M_B is a $r \times L$ sub-matrix of M, consists of rows $g_i(\forall i \in B)$. Since $g_j(j \in B^c)$ is included in the row space of M_B, we can conclude that there always exist a unique β_j for each D_j such that $\beta_j \diamond M_B = g_j$. Note that since it has K_j in advance, D_j would recover q_j w.p. 1 once given $g_j * Q$ which is hold as $N \to \infty$. This completes the proof for d_I.

As a result, we can conclude that each receiver $D_j(\forall j \in [L])$ would recover demanded $k_j = \lfloor N \cdot I(W_j) \rfloor [bits]$ reliably through r-ICPC transmissions under corresponding conditions, and the whole system would achieve R_t

$$\begin{aligned}
R_t &= \frac{k_1 + k_2 + \cdots + k_L}{rN} \\
&\xrightarrow{N \to \infty} \frac{1}{r} \sum_{i=1}^{L} I(W_i)
\end{aligned}$$

as claimed in Theorem 5 and Corollary 2.

Example 5. *Let us consider following M_1 and M_2 in degraded structure.*

$$M_1 = \begin{bmatrix} 1 & b & 1 & 0 \\ a & 1 & 0 & 0 \\ \bar{a} & \bar{b} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad M_2 = \begin{bmatrix} 1 & b & 1 & 1 \\ a & 1 & 0 & 1 \\ \bar{a} & \bar{b} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

In both cases suppose S chooses $B = \{1, 2, 4\}$ as an IC option. In M_1, D_3 would reliably decode ICPC words, hence B is feasible. However in M_2, D_3 fails in SC-decoding since it has no information on q_4 (or m_4), thus B is not a feasible solution. For M_2, $B = \{2, 3, 4\}$ is a feasible set that satisfies Condition 1 .

Note that when S builds M from received K_j ($\forall j \in [L]$) to figure out a feasible B, it can use only a subset $K'_j \subseteq K_j$ in order to reduce the rank of M. This procedure does not harm the decoding reliability since those ignored messages are still exist in D_j. Therefore, $g_j(i) = 0$ does not necessarily mean $K_j(i) = 0$ since it might be intentionally ignored in creating M.

4.4.5 The ICPC scheme

4.4.5.1 Encoding

Given $W_j(\forall j \in [L])$, each m_j is chunked into q_j. First, S builds a $L \times N$ matrix Q whose rows are denoted as q'_j such that $q_j \mapsto q'_j(A_j)$ where A_j is a information index set under W_j. Second, with reported K_1^L, it figures out a feasible B and a corresponding IC option

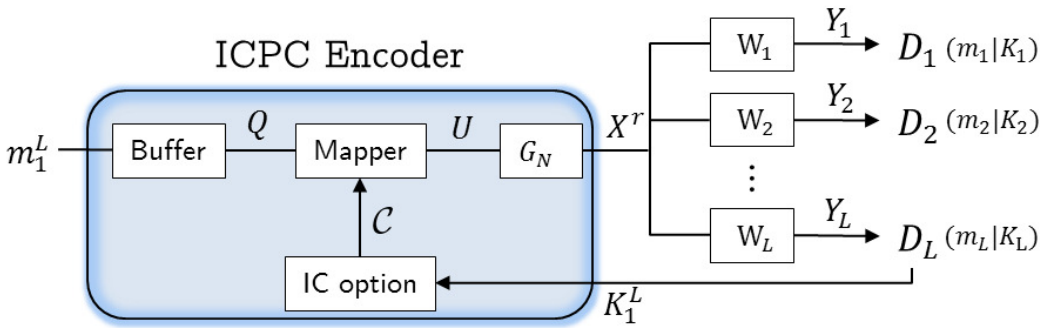

Figure 4.8 The ICPC system model where $U = \{u^N(i)|\forall i \in [r]\}$, $X^r = \{x^N(i) = u^N(i)G_N|\forall i \in [r]\}$ and $Y_j = \{y_j^N(i)|\forall i \in [r]\}$

$\mathcal{C} = (c_1, c_2, \cdots, c_r)$ as depicted in Fig. 4.8. Third, S chooses feasible r-linearly independent rows of M, $\alpha_j(\forall j \in [r])$ as index encoding vectors, based on $\mathcal{C}$, which is same as g_B such that $\alpha_j = g_{B(j)}$.

Then the η^{th} message for PC $u^N(\eta)$ which is consist of $u_{A_{\{\eta\}}}$ and $u_{F_{\{\eta\}}}$, is a linear combination of rows of Q by using α_η as a coefficient vector:

$$u^N(\eta) = \sum_{j=1}^{L} \alpha_{\eta j} \cdot q'_j \tag{4.37}$$

denoted as $u^N(\eta) = \alpha_\eta * Q$.

Finally, the message vector $u_{A_{\{\eta\}}}$ is encoded with a generator matrix G_N: $x^N(\eta) = u_{A_{\{\eta\}}} G_{A_{\{\eta\}}} \oplus u_{F_{\{\eta\}}} G_{F_{\{\eta\}}}$.

4.4.5.2 Decoding

Each D_j receives $Y_j = \{y_j^N(1) \cdots y_j^N(r)\}$ and recovers $\hat{q}_j$ with a decoder $d_j : Y_j \mapsto \hat{q}_j$ which is a two-tiered procedure of (d_S, d_I). The first tier $d_S : Y_j \mapsto \hat{U}_j$ is a successive

cancellation (SC) decoder given (Y_j, A_j, u_{F_j}) where $\hat{U}_j \subseteq U$ that is required to decode q_j. The second is the IC decoding that maps $\hat{U}_j$ to $\hat{q}_j$. In the proof of Theorem 5, it is shown that once $\hat{U}_j = U_j$ then $\hat{q}_j = q_j$. Since the index encoding and the decoding operations are deterministic linear operations, the ICPC error rate $P_e \to 0$ as $N \to \infty$.

If all of each g_j ($\forall j \in B^c$) are repetitions of g_i ($\forall i \in B$) the problem becomes trivial: Such D_j would decode q_j once receiving U_j reliably. The complete SI graph is a special case where $g_j = 1_L$ for $\forall j \in [L]$. Every D_j would reliably decodes q_j since $u^N = \sum_{j=1}^{L} q'_j$ and $K_j = m_{[L]\backslash j}$. In this case the $R_t = \frac{1}{N}\sum_{i=1}^{L} I(W_i)$ achievable ICPC scheme exists for all L.

Algorithm 8 ICPC Encoding

1: Given K_1^L, find feasible $\mathcal{C} = \{c_i | \forall i \in [r]\}$
2: Calculate A_j for $\forall j \in [L]$ and construct $Q : m_j \mapsto q'_j$
3: **for** $\eta = 1 : r$
4: **do** $u^N(\eta) = \sum_{j=1}^{L} \alpha_{\eta j} \cdot q'_j$
5: $x^N(\eta) = u_{A_{\{\eta\}}} \cdot G_{A_{\{\eta\}}} \oplus u_{F_{\{\eta\}}} \cdot G_{F_{\{\eta\}}}$
6: **end for**

Algorithm 9 ICPC Decoding

At D_j, $d_j(Y_j, A_j, u_{F_j})$ = $[d_S(N, A_j, u_{F_j}), d_I(U_j)]$
1: Create frozen bits from SI: $K_j \mapsto u_{F_j}$
2: SC-decoder $d_S(N, A_j, u_{F_j}) : Y_j \mapsto \hat{U}_j$
3: Index decoder $d_I : \hat{U}_j \mapsto \hat{q}_j$

4.4.6 Example: Partially Perfect Graph

Let us consider the simplest case that M consist of partially fully connected matrices such as

$$M = \begin{bmatrix} 1_{a_1} & & & 0 \\ & 1_{a_2} & & \\ & & \ddots & \\ 0 & & & 1_{a_r} \end{bmatrix}$$

$$= I_r \otimes (1_{a_1}, 1_{a_2}, \cdots, 1_{a_r})$$

where 1_{a_j} is $a_j \times a_j$ matrix whose elements are all ones, by row changes without performance loss, and $\otimes$ is kronecker product.

The column indices in 1_{a_j} determine the j-th coding solution c_j. Since all the sub-matrices form fully connected digraphs,

$$c_j = \bigoplus_{i=|a_{j-1}|+1}^{|a_j|} m_i, \quad \text{for} \quad \forall j \in [1, r] \tag{4.38}$$

Therefore, we can exploit the proposed ICPC scheme in Section 4.3 r-times, and the ICPC rate R_t would be

$$\begin{aligned} R_t &= \frac{k_1 + k_2 + \cdots + k_L}{rN} \\ &= \frac{1}{r} \sum_{i=1}^{L} \frac{k_i}{N} \\ &\xrightarrow{N \to \infty} \frac{1}{r} \sum_{i=1}^{L} I(W_i) \end{aligned}$$

Note that the Polar codeword length $N \to \infty$ constraint in the last equation, as assumed in Proposition 5. The fully connected SI case in Section 4.3 is one special case such that $r = 1$. Note that this R_t is optimal by considering two features:

1. $\sum_{j=1}^{L} I(W_j)$ is equivalent to the $cut - capacity$ of BBC which is the upper bound of the channel code's performance.

2. The index code length r is known as the optimal (shortest) length in scalar index coding systems.

4.5 ICPC for Probabilistic Side Information

Suppose that there are no deterministic SI feedback from the set of receiver, rather are probabilities, then M_P has the following format.

$$M_P = \begin{bmatrix} 1 & p_{12} & p_{13} & \cdots & p_{1L} \\ p_{21} & 1 & p_{23} & \cdots & p_{2L} \\ p_{31} & p_{32} & 1 & \cdots & p_{12} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ p_{L1} & p_{L2} & p_{L2} & \cdots & 1 \end{bmatrix} \tag{4.39}$$

where each $p_{ij} = P(m_j \in K_i)$ for D_i (or equivalently $P(m_{ij} = 1)$ in real SI matrix M). We assume that probabilities in this matrix could be acquired from various estimation operations at the source S. If $p_{ij} = p_{kl} = p$ for all pairs, it can be represented as a random graph $G(L, p)$.

There are few criteria that S can exploit.

1. Random ICPC scheme
2. Choose a candidate M from M_P that maximizes the expected achievable rate of linear ICPC scheme
3. Threshold based estimation of M (linear ICPC)

4.5.1 Random ICPC for non-identical B-DMCs

The most important advantage of the random IC is that S do not need to care about SI of every receivers, and it is suitable for broadcasting all the message to all receivers.

If $W_i = W_j$ for $\forall i, j \in [1, L]$ then the random ICPC scheme is simply a concatenation of random IC encoder and the PC encoder. Since $A_i = A_j$ for $\forall i, j \in [1, L]$ under this case, first randomly index coded words are divided into $k = |A_i|$ bits and they are mapped to the information bits of the PC encoder.

However, for non-identical channels $W_i \neq W_j$ for some $i, j \in [1, L]$ there exists a Random Linear ICPC (RLICPC) only with full SI (or equivalently with complete graphs). It can be easily verified using the Fig. 4.6 where in RLICPC, all coefficients α_{ij} are chosen randomly in $GF(2)$ and stored in the header. In case of full SI, each receiver D_j can recover q_j (when D_j demands only q_j) for $\forall j \in [1, L]$, however, suppose that D_2 does not have m_3 or m_4 in SI, K_2, the SC-dec would become unreliable, since it should have $u_{A_3 \setminus A_2}$ and $u_{A_4 \setminus A_3}$ for complete knowledge of u_{F_2}.

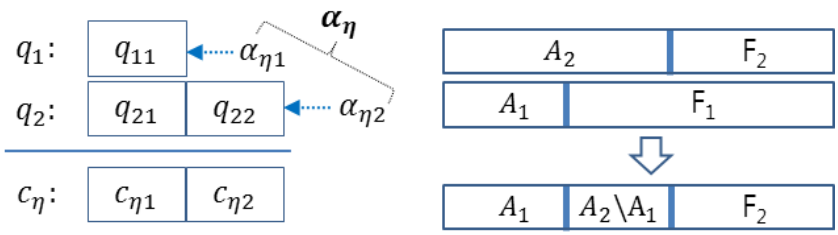

Figure 4.9 [L=2 RLICPC] If D_1 does not have m_2 in SI, K_1, the SC-dec would become unreliable. It should have $u_{A_2 \backslash A_1}$ for complete knowledge of u_{F_1}. Note that $\alpha_\eta \in GF(2)^2$ and chosen randomly.

To guarantee the reliable SC-dec S should limit the channel as the worst one (W_1 for degraded structure), which incur the negative bottleneck effect by decrease the achievable rate. The achievable rate in this case is

$$\begin{aligned} R &= \frac{k_1}{N} \\ &\xrightarrow{N\to\infty} I(W_1) \end{aligned} \tag{4.40}$$

which is smaller than the optimal R_o:

$$\begin{aligned} R_o &= \frac{\sum_{j=1}^{L} k_j}{LN} \\ &\xrightarrow{N\to\infty} \frac{1}{L}\left(\sum_{j=1}^{L} I(W_j)\right) \end{aligned} \tag{4.41}$$

Note that these rates is for the case where S delivers all messages to every receivers, that is different to the previous model.

4.5.2 Expected rate maximization

As another criterion, we can choose one candidate IC solution for M that maximizes the expected achievable rate of linear ICPC scheme. This method seems quite reasonable in exploiting the average rate, however the major constraint is the computational burden: for L-Rx, S try to consider $2^{L(L-1)}$ terms whether each one is 0 or 1. Thus the whole case of numbers $2^{2^{L(L-1)}}$ would diverge beyond a computational capability even for moderate L.

In this part we investigate finite case of $L = 2$ model and propose a condition where IC gain exist. There are 4 possible M and for each case we calculate the expected rate $\bar{R}_t$. For simplicity of notation, let $p_{12} = \alpha, p_{21} = \beta$.

1. $(m_{12}, m_{21}) = (0, 0)$

   In this case, S assume that $M = \begin{bmatrix} 1 & 0 \\ & \\ 0 & 1 \end{bmatrix}$, hence $r = 2$ and $\mathcal{C} = (c_1, c_2)$ for $c_1 = q_1, c_2 = q_2$ where each queue length $|q_j| = k_j$. Then, the deterministic achievable rate (since no IC is applied) is:

   $$R_t = \frac{k_1 + k_2}{2N} \tag{4.42}$$

   For non-degraded case,

   $$R_t = \frac{\alpha k_1 + \beta k_2}{2N} \tag{4.43}$$

2. $(m_{12}, m_{21}) = (1, 0)$

   In this case, S assume that $M = \begin{bmatrix} 1 & 1 \\ & \\ 0 & 1 \end{bmatrix}$, Still $r = 2$ and $\mathcal{C} = (c_1, c_2)$ for $c_1 =$

$f(q_1, q_2), c_2 = q_2$ where $f(\cdot)$ denote the operation for the ICPC. Then,

$$\bar{R}_t = \frac{\alpha k_1 + \beta k_2 + \bar{\beta}(k_2 - k_1) + k_2}{2N} \tag{4.44}$$

For non-degraded case,

$$\bar{R}_t = \frac{\alpha k_1 + \beta k_2 + k_2}{2N} \tag{4.45}$$

3. $(m_{12}, m_{21}) = (0, 1)$ In this case, S assume that $M = \begin{bmatrix} 1 & 0 \\ & \\ 1 & 1 \end{bmatrix}$, Still $r = 2$ and $\mathcal{C} = (c_1, c_2)$ for $c_1 = q_1, c_2 = f(q_1, q_2)$ where $f(\cdot)$ denote the operation for the ICPC.

   Then,

$$\bar{R}_t = \frac{k_1 + \alpha k_1 + \beta k_2 + \bar{\beta}(k_2 - k_1)}{2N} \tag{4.46}$$

   For non-degraded case,

$$\bar{R}_t = \frac{k_1 + \alpha k_1 + \beta k_2}{2N} \tag{4.47}$$

4. $(m_{12}, m_{21}) = (1, 1)$ In this case, S assume that $M = \begin{bmatrix} 1 & 1 \\ & \\ 1 & 1 \end{bmatrix}$, Still $r = 1$ and $\mathcal{C} = (c)$ for $c = f(q_1, q_2)$ where $f(\cdot)$ denote the operation for the ICPC. Then,

$$\bar{R}_t = \frac{\alpha k_1 + \beta k_2 + \bar{\beta}(k_2 - k_1)}{N} \tag{4.48}$$

   For non-degraded case,

$$\bar{R}_t = \frac{\alpha k_1 + \beta k_2}{N} \tag{4.49}$$

These results are summarized below.

(m_{12}, m_{21})	Degraded	Non-degraded
(0,0)	$\frac{k_1+k_2}{2N}$	$\frac{k_1+k_2}{2N}$
(1,0)	$\frac{\alpha k_1+\beta k_2+\bar{\beta}(k_2-k_1)+k_2}{2N}$	$\frac{\alpha k_1+\beta k_2+k_2}{2N}$
(0,1)	$\frac{\alpha k_1+\beta k_2+\bar{\beta}(k_2-k_1)+k_1}{2N}$	$\frac{\alpha k_1+\beta k_2+k_1}{2N}$
(1,1)	$\frac{\alpha k_1+\beta k_2+\bar{\beta}(k_2-k_1)}{N}$	$\frac{\alpha k_1+\beta k_2}{N}$

After calculation of these expected rates, S chooses one of those candidate M which would achieve the highest rate in average among these four. For example, $(m_{12}, m_{21}) = (1, 1)$ case will be chosen under degraded structure, if its expected rate is the highest. Therefore it will be picked if

$$\alpha k_1 + \beta k_2 + \bar{\beta}(k_2 - k_1) > k_2 \tag{4.50}$$

or equivalently, if $\alpha + \beta > 1$. In the non-degraded case, without loss of generality assume $k_1 \leq k_2$. Then $(m_{12}, m_{21}) = (1, 1)$ case will be chosen if

$$\alpha k_1 + \beta k_2 > k_2 \tag{4.51}$$

or equivalently, if $\frac{\alpha}{1-\beta} > \frac{k_2}{k_1}$.

Unfortunately, for $L \geq 3$, calculations of every case of numbers of M would be almost impossible for even moderate L. To solve this issue we need another approach. At least we can calculate the expected rate when S transmits rank one IC solution which corresponds to

the complete graph.

$$\bar{R}_t = \frac{1}{N}\left(\sum_{i=1}^{L}\left[\bar{p}^{i-1}p^{L-i}\sum_{j=1}^{i}(k_i - k_{j-1})\right]\right) \tag{4.52}$$

where $k_0 = 0$.

4.5.3 Expected achievable rate via Random graph

If all the probabilities are identical to p, the SI graph can be represented by a random digraph $\bar{G}(L,p)$. In [54], the authors found the lower bound of of the minimum IC length of $undirected$ graph $G(L,p)$:

$$minrk_{\mathbb{F}}(G(L,p)) = \Omega(\sqrt{L}) \tag{4.53}$$

where $\mathbb{F}$ is its field. They proved that there exists some constant p' such that the minimum IC length for $\bar{G}(L,p)$ is the same as the one for $G(L,p')$. Hence for digraph,

$$minrk_{\mathbb{F}}(\bar{G}(L,p)) = \omega(\sqrt{L}) \tag{4.54}$$

However, these lower bounds $\Omega(\sqrt{L})$ and $\omega(\sqrt{L})$ are asymptotically achieved as $L \to \infty$.

Therefore, for large L, the following lemma holds asymptotically.

Lemma 7. *For any $\bar{G}(L,p)$, if there exist at least one feasible IC solutions, then the expected achievable rate would be $\bar{R}_t = o(\sqrt{L})$*

Proof. The proof of Lemma 7 follows from the proof of Theorem 5. If a chosen IC solution is feasible, the source S would transfer every demands from receivers; D_j would receive

$k_j[bits]$ for $\forall j \in [1, L]$ reliably as $N \to \infty$ and $L \to \infty$ within $\omega(\sqrt{L})$-transmissions of ICPC words. Therefore it holds

$$\bar{R}_t = \frac{1}{\omega(\sqrt{L})} \sum_{j=1}^{L} I(W_j) \tag{4.55}$$

where $\frac{1}{\omega(\sqrt{L})}$ is equivalent to $o(L^{-\frac{1}{2}})$. Since $\sum_{j=1}^{L} I(W_j) \leq L$ which means $\sum_{j=1}^{L} I(W_j) \in o(L)$, then we can conclude that $\bar{R}_t \in o(\sqrt{L})$ which complete the proof. □

Let us change the system model. Assume a storage system where there is single server and multiple storages as we have used until now. The marginal channels in this bidirectional broadcast channel are degraded in that $W_i \preceq W_{i+1}$ for $\forall i \in [L-1]$. In the system, the controller (server) S has to maintain all storages D_j to have equal message set over their corresponding marginal channels W_j.

$$K_j(n_o + r_N) = \cup_{\forall i \in [L]} K_i(n_o) \tag{4.56}$$

Kolchin [56] proved the probability mass function of the rank of the binary random matrices

$$\lim_{L \to \infty} P(r) = 2^{-(L-r)^2} \prod_{i=L-r+1}^{\infty} (1 - \frac{1}{2^i}) \prod_{i=1}^{L-r} (1 - \frac{1}{2^i})^{-1} \tag{4.57}$$

Therefore, from (4.57), we can claim that is there exist any feasible IC options, the average achievable rate would be

$$\bar{R}_t = \frac{1}{E[r]} \sum_{j=1}^{L} I(W_j)$$

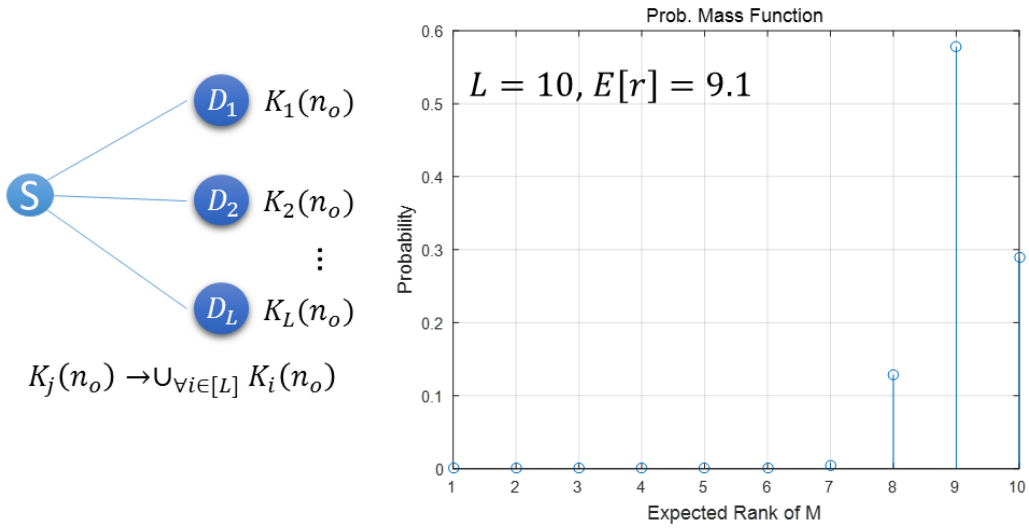

Figure 4.10 The expected rank of binary random matrices

4.6 Summary

In this chapter, we developed the joint coding scheme of nested Polar codes and the index codes which is denoted as ICPC schemes and proved that via ICPC the system can achieve the rate of $\frac{1}{r}\sum_{j=1}^{L} I(W_j)$.

In full SI case, I proved that there exist ICPC schemes w.p.1, and for arbitrary SI patterns we showed in Theorem 5 that ICPC schemes that achieve such rate would exist only when there are at least one feasible IC solution for each SI pattern and M. In addition, we also consider probabilistic SI where those information might be feedbacked from receivers or S estimates existence of messages in each receiver. We model the SI graph as a random digraph $\bar{G}(L, p)$ where edge connection probabilities are all identical to p, and suggest the upperbound of the expected rate would be $o(\sqrt{L})$ when there exist at least one feasible IC

solution for ICPC shcemes, exploiting the minimum rank of a random graph result in the previous literature and Theorem 5.

Chapter 5

Conclusions

In Part I, we proved that for deterministic CPs in non-identical channel models, polar codes can achieve the sample mean of bit channel capacities and extended the PC scheme to random channel parameter case. The key contribution is a new system model where the transmitter and the receiver knows only the channel parameter distribution instead of channel parameter itself. The existence can be proved by the average sense on symmetric capacity I. One may use an inverse function I^{-1} (or an approximate version) or pre-calculated table look-up for the other cases of channels. However, if the underlying channel type is BEC, the coding scheme can become simpler. Note that for a BEC with crossover probability ϵ, its symmetric capacity I is the affine function of ϵ. Then, we have the relation $E[I(\varepsilon)] = I(\overline{\epsilon})$ where $\overline{\epsilon}$ is the expectation of the random variable $\varepsilon \sim f_{\varepsilon}(\epsilon)$.

By applying multiple streams of polar codewords, we prove that the average capacity of any B-DMCs under our scenarios is achievable. However, this is obtained by sacrificing

the latency and complexity, since they stack multiple blocks during encoding and decoding process. Hence, these schemes might not be suitable in the systems where low latency or low complexity is required. Rather, it is more practical in storage systems such as flash memory devices where throughput is much important than latency. Especially, for flash memories, statistical responses such as a voltage threshold would change with time and with the number of accesses to a cell block. Hence, as the storage capacity increases, it is inefficient for a storage controller, to figure out exact states of every blocks or cells. If statistics on their changes are given instead, we can manage cells more efficiently using the proposed polar coding scheme. In addition, in the case of parallel channels where there exist statistically different random disturbances across channels, it would be difficult to track all the channel parameters. However, if their statistics are known to the transmitter and the receiver, we can deliver data up to the average capacity through polar codes by sacrificing latency. In such cases, polar codes are a promising option which maximizes the throughput.

Under the non-independent channel scenario, we assume that N transmit channels are grouped into channels with size r which is a power of two, so that we can deal with the scenario as a non-binary system. If N is not divisible by r ($N \mod r \neq 0$), puncturing may be used to fit the system into a q-ary system. The proposed polar codes appear to be promising for applications where only the knowledge of channel parameter distribution is available, and can be practical for storage applications such as flash memory devices.

We also consider the importance of the channel interleaver that could enhance the

system reliability. The heuristic algorithm, Peeler method is proposed and it shows better convergence to the capacity compared to random permutations.

In Part II, we developed the joint coding scheme of nested Polar codes and the index codes which is denoted as ICPC schemes and proved that via ICPC the system can achieve the rate of $\frac{1}{r}\sum_{j=1}^{L} I(W_j)$.

In full SI case, I proved that there exist ICPC schemes w.p.1, and for arbitrary SI patterns we showed in Theorem 5 that ICPC schemes that achieve such rate would exist only when there are at least one feasible IC solution for each SI pattern and M. In addition, we also consider probabilistic SI where those information might be feedbacked from receivers or S estimates existence of messages in each receiver. We model the SI graph as a random digraph $\bar{G}(L,p)$ where edge connection probabilities are all identical to p, and suggest the upperbound of the expected rate would be $o(\sqrt{L})$ when there exist at least one feasible IC solution for ICPC shcemes, exploiting the minimum rank of a random graph result in the previous literature and Theorem 5.

However, it is still open questions whether there are universal ICPC schemes that can achieve the capacity for a random graph $\bar{G}(L,p)$, and that for a random graph with non-identical probabilities.

Appendix A

A.1 Proof of (2.25)

Proof. We can prove (2.25) simply by applying the arithmetic-geometric mean inequality on $Z(W_2^{(1)})$. Let us review the development process of (2.28):

$$\begin{aligned}
&Z(W_2^{(1)}) \\
&= \sum_{y_1^2} \sqrt{W_2^{(1)}(y_1^2|u_1=0) W_2^{(1)}(y_1^2|u_1=1)} \\
&= \sum_{y_1^2} \sqrt{\sum_{u_2} \frac{1}{2} W_{(1)}(y_1|u_2) W_{(2)}(y_2|u_2)} \\
&\qquad \cdot \sqrt{\sum_{u_2'} \frac{1}{2} W_{(1)}(y_1|1+u_2') W_{(2)}(y_2|u_2')} \\
&= \sum_{y_1^2} \frac{1}{2} \sqrt{W_{(1)}(y_1|0) W_{(2)}(y_2|0) + W_{(1)}(y_1|1) W_{(2)}(y_2|1)} \\
&\qquad \cdot \sqrt{W_{(1)}(y_1|1) W_{(2)}(y_2|0) + W_{(1)}(y_1|0) W_{(2)}(y_2|1)}
\end{aligned} \tag{A.1}$$

Define shorthand notations of $A = W_{(1)}(y_1|0), B = W_{(1)}(y_1|1), C = W_{(2)}(y_2|0)$, and

$D = W_{(2)}(y_2|1)$, we can rewrite above equation as follows

$$\begin{aligned}
Z(W_2^{(1)}) &= \sum_{y_1^2} \frac{1}{2}\sqrt{ABC^2 + CDA^2 + CDB^2 + ABD^2} \\
&\geq \sum_{y_1^2} \frac{1}{2}\sqrt{4 \cdot ABCD} \\
&= \sum_{y_1} \sqrt{AB} \sum_{y_2} \sqrt{CD} \\
&= \sum_{y_1} \sqrt{W_{(1)}(y_1|0)W_{(1)}(y_1|1)} \\
&\cdot \sum_{y_2} \sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)} \\
&= Z(W_{(1)})Z(W_{(2)}) \\
&= Z(W_2^{(2)})
\end{aligned} \tag{A.2}$$

which the inequality is from the arithmetic and geometric mean relation. □

A.2 Proof of (2.36)

Proof. The proof of this equation is straightforward.

$$\begin{aligned}
&Z(W_2^{(1)}) \\
&= \sum_{y_1^2} \sqrt{W_2^{(1)}(y_1^2|u_1 = 0)W_2^{(1)}(y_1^2|u_1 = 1)} \\
&= \sum_{y_1^2} \sqrt{\sum_{u_2} \frac{1}{2} W_{(1)}(y_1|u_2)W_{(2)}(y_2|u_2)}
\end{aligned}$$

$$\cdot \sqrt{\sum_{u_2'} \frac{1}{2} W_{(1)}(y_1|1+u_2')W_{(2)}(y_2|u_2')}$$

$$= \sum_{y_1^2} \frac{1}{2} \sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|0) + W_{(1)}(y_1|1)W_{(2)}(y_2|1)}$$

$$\cdot \sqrt{W_{(1)}(y_1|1)W_{(2)}(y_2|0) + W_{(1)}(y_1|0)W_{(2)}(y_2|1)}$$

$$\leq \sum_{y_1^2} \frac{1}{2} [\sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|0)} + \sqrt{W_{(1)}(y_1|1)W_{(2)}(y_2|1)}]$$

$$\cdot [\sqrt{W_{(1)}(y_1|1)W_{(2)}(y_2|0)} + \sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|1)}]$$

$$- \sum_{y_1^2} \sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|0)W_{(1)}(y_1|1)W_{(2)}(y_2|1)}$$

after calculations, it becomes

$$\begin{aligned} &\frac{1}{2} \sum_{y_1^2} \Big(W_{(1)}(y_1|0) \sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)} \\ &\quad + W_{(1)}(y_1|0) \sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)} \\ &\quad + W_{(1)}(y_1|0) \sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)} \\ &\quad + W_{(1)}(y_1|0) \sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)} \Big) \\ &\quad - \sum_{y_1^2} \sqrt{W_{(1)}(y_1|0)W_{(2)}(y_2|0)W_{(1)}(y_1|1)W_{(2)}(y_2|1)} \end{aligned} \tag{A.3}$$

$$= Z(W_{(1)}) + Z(W_{(2)}) - Z(W_{(1)})Z(W_{(2)})$$

Therefore, $Z(W_2^{(1)}) \leq Z(W_{(1)}) + Z(W_{(2)}) - Z(W_{(1)})Z(W_{(2)})$ is satisfied for any binary input channel parameters. □

A.3 Proof of (2.37)

Proof. The proof of this equation is straightforward.

$$
\begin{aligned}
& Z(W_2^{(2)}) \\
&= \sum_{y_1^2,u_1} \sqrt{W_2^{(2)}(y_1^2,u_1|u_2=0)W_2^{(2)}(y_1^2,u_1|u_2=1)} \\
&= \sum_{y_1^2,u_1} \frac{1}{2}\sqrt{W_{(1)}(y_1|u_1)W_{(2)}(y_2|0)} \\
&\cdot \sqrt{W_{(1)}(y_1|u_1+1)W_{(2)}(y_2|1)} \\
&= \sum_{y_2} \sqrt{W_{(2)}(y_2|0)W_{(2)}(y_2|1)} \\
&\cdot \sum_{y_1,u_1} \frac{1}{2}\sqrt{W_{(1)}(y_1|u_1)W_{(1)}(y_1|u_1+1)} \\
&= Z(W_{(2)})Z(W_{(1)})
\end{aligned}
\tag{A.4}
$$

□

The fourth equation comes from that the summation over u_1 is a sum of two same terms. In addition, the following relation hold with the aid of (2.24):

$$
Z(W_2^{(2)}) \leq \min(Z(W_{(1)}), Z(W_{(2)})) \tag{A.5}
$$

It can be verified simply by subtracting either of right terms from the left term.

A.4 Proof of the number of equivalent channel combinations

According to the basic theory of Polar codes, channels are recursively evolved such that $(W_N^{(i)}, W_N^{(i)}) \mapsto (W_{2N}^{(2i-1)}, W_{2N}^{(2i)})$ for some set of binary input discrete memoryless channels. We denote each mapping as follows:

$$\mathcal{F}_{n+1} : W_N^{(i)}, W_N^{(i)} \mapsto W_{2N}^{(2i-1)} \tag{A.6}$$

$$\mathcal{G}_{n+1} : W_N^{(i)}, W_N^{(i)} \mapsto W_{2N}^{(2i)} \tag{A.7}$$

and it was proved that polarizations would occur in non-identically distributed channels.

Recall that functional $\mathcal{F}_n$ and $\mathcal{G}_n$ share the same set of functions as an input. Furthermore, those input functions are also results of functionals $\mathcal{F}_{n-1}$ or $\mathcal{G}_{n-1}$ due to the recursive structure of channel evolution.

Let us define an swapping operation $S : (\alpha, \beta) \mapsto (\beta, \alpha)$. Then we can easily check that $\mathcal{F}_n$ and $\mathcal{G}_n$ are invariant to S.

We also define a functional $\mathcal{H}_n$ which includes both $\mathcal{F}_n$ and $\mathcal{G}_n$. $\mathcal{H}_n$ can represent its members since $\mathcal{F}_n$ and $\mathcal{G}_n$ are equivalent in counting the number of cases, and hence, it is also invariant to S. In a similar way, we define $\mathcal{W}_{n+1}$ which includes $W_{2N}^{(2i-1)}$ and $W_{2N}^{(2i-1)}$.

Now we rewrite (A.6) and (A.7) using $\mathcal{H}$ and $\mathcal{W}$:

$$\mathcal{W}_{n+1} = \mathcal{H}_{n+1}(\mathcal{W}_n, \mathcal{W}_n). \tag{A.8}$$

Due to the recursive structure in channel evolutions, each $W_N^{(i)}$ is calculated again through $\mathcal{H}_n$. Then it is in general the form of

$$\mathcal{W}_{n+1} = \mathcal{H}_{n+1}(\mathcal{H}_n(\mathcal{W}_{n-1}, \mathcal{W}_{n-1}), \mathcal{H}_n(\mathcal{W}_{n-1}, \mathcal{W}_{n-1})). \tag{A.9}$$

Define the number of operations of $\mathcal{H}$ that is required to recursively reach $\mathcal{H}_n$ as χ_n. Then

$$\chi_{n+1} = 2\chi_n + 1 \tag{A.10}$$

By solving this recurrence formula, we can get $\chi_n = N - 1$. Recalling that $\mathcal{H}$ is invariant to S which indicates there are two number of cases for each $\mathcal{H}$, the number of combinations of the B-DMCs that would result in the same information set of Polar codes is 2^{N-1}.

Hence the number of representative combinations those which may have different information sets for length N parallel Polar coding systems is $\frac{N!}{2^{N-1}}$.

Bibliography

[1] E. Arikan, ”Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Transactions on Information Theory*, vol.55, no.7, pp.3051-3073, July 2009

[2] R. Mori, T. Tanaka, ”Performance of polar codes with the construction using density evolution,” *IEEE Communications Letters*, vol.13, no.7, pp.519-521, July 2009

[3] S. Boyd, L. Vandenberghe, ”Convex Optimization,” Cambridge University Press, 2004

[4] E. Hof, I. Sason, and S. Shamai, ”Polar coding for degraded and non-degraded parallel channels,” *IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI)*, pp.550-554, Nov. 2010

[5] E. Hof, I. Sason, S. Shamai, and Chao Tian, ”Capacity-achieving polar codes for arbitrarily permuted parallel channelsl,” *IEEE Transactions on Information Theory*, vol.59, no.3, pp.1505-1516, March 2013

[6] M. Alsan, ”Conditions for robustness of polar codes in the presence of channel mismatch,” ArXiv:1303.2379 [cs.IT], 2013.

[7] M. Alsan, E. Telatar, ”A simple proof of polarization and polarization for non-stationary

channels," *IEEE International Symposium on Information Theory (ISIT) 2014*, pp.301-305, June 2014

[8] Wolfram alpha symbolic equation calculator [Online]. Available: http://www.wolframalpha.com

[9] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *The Annals of Mathematical Statistics*, vol. 3, no. 4, pp.1229-1241, 1959.

[10] S. H. Hassani, S. B. Korada, and R. Urbanke, "The compound capacity of polar codes," *47th Annual Allerton Conference*, pp.16-21, Sept. 2009

[11] H. Mahdavifar, M. El-Khamy, Jungwon Lee, and Inyup Kang, "Compound polar codes," *Information Theory and Applications Workshop (ITA)*, pp.1-6, Feb. 2013

[12] E. Sasoglu, "Polar codes for discrete alphabets," *IEEE International Symposium on Information Theory Proceedings (ISIT) 2012*, pp.2137-2141, July 2012

[13] W. Park, A. Barg, "Polar codes for q-ary channels, q = 2^r," arXiv:1107.4965v3, Jan. 2012

[14] P. Billingsley, "Probability and Measure, Anniversary Edition", Wiley, 2012, pp. 487-495.

[15] R. G. Gallager, "Information Theory and Reliable Communication." New York: Wiley, 1968, pp. 524.

[16] A. Eslami and H. Pishro-Nik, "A practical approach to polar codes," *IEEE International Symposium on Information Theory Proceedings (ISIT) 2011*, pp.16-20, July 2011

[17] D. Tuninetti and C. Fragouli, "Processing along the way: forwarding vs. coding," *ISITA 2004*, Parma, Italy, Oct. 2004.

[18] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. on Info. Theory*, vol. 46, no. 4, pp. 1204-1216, July 2000.

[19] M. Charikar and A. Argawal "On the advantage of network coding for improving network throughput," *IEEE Info. Theory Workshop*, San Antonio, Oct. 2004.

[20] S.-Y. R. Li, R. W. Yeung, and N. Cai. "Linear network coding", *IEEE Trans. on Info. Theory*, vol. 49. no. 2, pp. 371-381, Feb. 2003.

[21] T. Ho, M. Medard, J. Shi, M. Effros, and D. R. Karger "On randomized network coding," *Proc. Allerton Conf. on Comm., Control, and Computing*, 2003.

[22] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the air: practical wireless network coding," *IEEE/ACM Trans. on Networking*, vol. 16. no. 3. pp. 497-510, June 2008.

[23] M. Ghaderi, D. Towsley, and J. Kurose, "Reliability gain of network coding in lossy wireless networks," *IEEE INFOCOM 2008*, Apr. 2008.

[24] Z. Fang and M. Medard "On analyzing and improving COPE performance," *Information Theory and Applications Workshop*, Feb. 2010.

[25] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman "A simple cooperative diversity method based on network path selection," *IEEE J. on Sel. Areas in Comm.*, vol. 24. no. 3, pp. 659-672, Mar. 2006.

[26] J. G. Proakis and M. Salehi, Digital Communication, 4th ed. McGrawHill, 2001.

[27] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," *Proc. Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, 2003.

[28] M. Xiao, M. Skoglund, "Design of network codes for multiple-user multiple-relay wireless networks," *ISIT 2009*, Seoul, Korea, June 2009.

[29] M. Xiao, M. Skoglund, "Multiple-User Cooperative Communications Based on Linear Network Coding," *IEEE Trans. on Comm*, vol. 58, no. 12, pp. 3345-3351, Dec. 2010.

[30] L. Lu, M. Xiao, L.K. Rasmussen, "Relay-Aided Broadcasting with Instantaneously Decodable Binary Network Codes," *ICCCN 2011*, July, 2011.

[31] S. E. Rouayheb, A. Sprintson and C. Georghiades, "On the Index Coding Problem and Its Relation to Network Coding and Matroid Theory," *IEEE Transactions on Information Theory*, vol.56, no.7, pp.3187-3195, July 2010

[32] E. Sasoglu, Polar Coding Theorems for Discrete Systems. Ph.D. Dissertation, EPFL, Lausanne, Switzerland, 2011.

[33] Aria G. Sahebi, S. Sandeep Pradhan, "Nested Polar Codes Achieve the Shannon Rate-Distortion Function and the Shannon Capacity," arXiv:1401.6482 [cs.IT], Jan 2014

[34] Y. Birk and T. Kol., "Informed-source coding-on-demand (ISCOD) over broadcast channels," *Proc. IEEE Conf. on Comput. Commun. (INFOCOM)*, pp. 1257–1264, 1998.

[35] Y. Birk and T. Kol., "Coding-on-demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients," *IEEE Transactions on*

Information Theory, vol. 52, no 6, pp. 2825–2830, 2006.

[36] Z. Bar-Yossef, Z. Birk, T. S. Jayram, and T. Kol. "Index coding with side information," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1479–1494, 2011.

[37] R. Peeters, "Orthogonal representations over finite fields and the chromatic number of graphs," Combinatorica, vol. 16, no. 3, pp. 417–431, 1996.

[38] S.H. Dau, V. Skachek, and Y. M. Chee, "Error Correction for Index Coding With Side Information,"*IEEE Transactions on Information Theory*, vol.59, no.3, pp.1517-1531, March 2013

[39] L. Xiao, T. E. Fuja, J. Kliewer, and D. J. Jr. Costello, "Nested codes with multiple interpretations," 40th Annual Conference on Information Sciences and Systems, pp. 851-856, March 2006

[40] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested Polar Codes for Wiretap and Relay Channels," Communications Letters, IEEE, vol.14, no.8, pp.752-754, August 2010

[41] M. Bloch, L. Luzzi and J. Kliewer, "Strong coordination with polar codes," Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on, pp.565-571, 2012

[42] R. Blasco-Serrano, R. Thobaben, M. Andersson, V. Rathi, and M. Skoglund, "Polar Codes for Cooperative Relaying," , IEEE Transactions on Communications, vol.60, no.11, pp.3263-3273, 2012

[43] S. Eghbalian-Arani and H. Behroozi, "Polar Codes for a Quadratic-Gaussian Wyner-Ziv Problem," Wireless Communication Systems (ISWCS 2013), Proceedings of the Tenth International Symposium on, Aug. 2013

[44] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, EPFL, 2009.

[45] K. Marton, "The capacity region of deterministic broadcast channels," in Prof. of the IEEE Intern. Symp. on Info. Theory, (Paris-Cachan), 1977.

[46] M. S. Pinsker, "Capacity of noiseless broadcast channels," Probl. Inform. Transm., pp. 97–102, June 1978.

[47] M. Andersson, F. Schaefer, T. Oechtering, M. Skoglund, "Polar Coding for Bidirectional Broadcast Channels with Common and Confidential Messages," *IEEE Journal onSelected Areas in Communications*, vol.31, no.9, pp.1901-1908, 2013

[48] P. Maymounkov, A. Harvey, and D. S. Lun, "Methods for Efficient Network Coding," in Proc. 44th Annual Allerton Conference on Communication, Control, and Computing, 2006.

[49] P. A. Chou, Y. Wu, and K. Jain, "Practical Network Coding," in Proc. Allerton Conference on Communication, Control, and Computing, 2003.

[50] A. Heidarzadeh, A. H. Banihashemi, "Overlapped Chunked network coding," Information Theory Workshop (ITW), Jan. 2010

[51] A. Heidarzadeh, A. H. Banihashemi, "Analysis of overlapped chunked codes with

small chunks over line networks," IEEE International Symposium on Information Theory Proceedings (ISIT), pp.801-805, July 2011

[52] Bin Tang, Shenghao Yang, Yitong Yin, Baoliu Ye, and Sanglu Lu, "Expander graph based overlapped chunked codes," IEEE International Symposium on Information Theory Proceedings (ISIT), pp.2451-2455, July 2012

[53] G. Joshi, E. Soljanin, "Round-robin overlapping generations coding for fast content download," IEEE International Symposium on Information Theory Proceedings (ISIT), pp.2740-2744, July 2013

[54] I. Haviv, M. Langberg , "On linear index coding for random graphs," arXiv:1107.0390H, 2011

[55] I. Haviv, M. Langberg, "On linear index coding for random graphs," IEEE International Symposium on Information Theory (ISIT), pp.2231-2235, 2012

[56] V. F. Kolchin, Random Graphs, Cambridge University Press, 252 pg., 1999.

한글초록

본 논문은 Part I의 비동형 대칭 이진 이산 무기억 채널에서 채널 용량을 달성 하는 폴라 부호의 설계 기법 및 증명과 Part II의 비동형 이진 독립 무기억 채널에서 인덱스 부호와 폴라 부호의 연계기법을 통한 최적의 전송률을 달성하는 연계 기법에 대한 설계로 구성된다.

Part I에서는 먼저 각 채널의 통계적 특성을 대변하는 채널 파라미터가 결정적인 형태로 부호기와 복호기에 주어지는 경우대 대해 다루며, 두번째로 이 파라미터들이 결정적이 아닌 랜덤한 값으로써 주어지는 경우에 대하여 적합한 폴라 부호 기법에 대해 기술한다. 후자는 다시 두가지의 하위 경우로 나뉘는데 하나는 모든 파라미터들이 단 하나의 확률 분포에 대한 실현값인 경우이고, 또다른 한가지는 각 파라미터들이 각각의 서로 다른 확률 분포의 실현값인 경우이다. 폴라 부호를 이용하여 결정적인 경우와 랜덤한 실현값으로 주이지는 모든 경우에 대하여 평균 채널 용량을 달성 할 수있음을 증명한다.

이에 더해 결정적 채널 파라미터가 가정된 시스템에서 채널 입력으로 사용되는 정보 벡터의 치환 연산의 중요성에 대하여 논한다. 적절한 치환 연산을 이론적 상한 값인 채널용량에 대한 수렴속도를 향상 시킬수 있음을 예시를 통해 보이고 휴리스틱 치환 알고리즘을 개발하여 달성 전송률 또는 시스템 신뢰도를 향상 시킬수 있음을 보인다.

Part II에서는 폴라 부호와 인덱스 부호를 접합시켜 일종의 연계된 소스-채널 부호 설계 기법을 개발하고 제안된 기법이 최적의 전송률을 달성함을 보인다. 먼저 인덱스 부호에서 수신노드에서 송신노드로 전달되는 부가정보를 통해 그려지는 그래프가 완전그래프일때 항상 최적의 달성 기법이 존재함을 보이고, 이를 임의의 부가정보 패턴이 주어지는 경우로 확장한다. 완전 그래프가 그려지는 경우와 달리 임의의 패턴으로 주어지는 경우는 부가정보들이 특정 조건을 만족하는 경우에 한하여 최적 전송률을 달성하게됨을 보이고 이를 만족하는 인덱스-폴라 부호 설계 기법을 제안한다. 마지막으로 부가정보가 결정적으로 주어지지 않고 존재성을 표현하는 확률로써 주어지는 경우 제안된 연계기법을 이용한 평균 전송률에 대하여 논한다.

주요어: 폴라 부호, 비동형 이산 무기억 채널, 인덱스 부호, 인덱스-폴라 부호 연계 기법

학번: 2010-30978